NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM REPORT 29

DIGITAL-COMPUTER-CONTROLLED TRAFFIC SIGNAL SYSTEM FOR A SMALL CITY

HIGHWAY RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING

HIGHWAY RESEARCH BOARD 1966

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM
REPORT 29

DIGITAL-COMPUTER-CONTROLLED TRAFFIC SIGNAL SYSTEM FOR A SMALL CITY

MORTON I. WEINBERG, HARVEY GOLDSTEIN, TERENCE J. McDADE,
AND ROBERT H. WAHLEN
CORNELL AERONAUTICAL LABORATORY
BUFFALO, NEW YORK

RESEARCH SPONSORED BY THE AMERICAN ASSOCIATION
OF STATE HIGHWAY OFFICIALS IN COOPERATION
WITH THE BUREAU OF PUBLIC ROADS

SUBJECT CLASSIFICATION:
TRAFFIC CONTROL AND OPERATIONS
TRAFFIC FLOW
TRAFFIC MEASUREMENTS

HIGHWAY RESEARCH BOARD
DIVISION OF ENGINEERING NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING 1966

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

Systematic, well-designed research provides the most effective approach to the solution of many problems facing highway administrators and engineers. Often, highway problems are of local interest and can best be studied by highway departments individually or in cooperation with their state universities and others. However, the accelerating growth of highway transportation develops increasingly complex problems of wide interest to highway authorities. These problems are best studied through a coordinated program of cooperative research.

In recognition of these needs, the highway administrators of the American Association of State Highway Officials initiated in 1962 an objective national highway research program employing modern scientific techniques. This program is supported on a continuing basis by Highway Planning and Research funds from participating member states of the Association and it receives the full cooperation and support of the Bureau of Public Roads, United States Department of Commerce.

The Highway Research Board of the National Academy of Sciences-National Research Council was requested by the Association to administer the research program because of the Board's recognized objectivity and understanding of modern research practices. The Board is uniquely suited for this purpose as: it maintains an extensive committee structure from which authorities on any highway transportation subject may be drawn; it possesses avenues of communications and cooperation with federal, state, and local governmental agencies, universities, and industry; its relationship to its parent organization, the National Academy of Sciences, a private, non-profit institution, is an insurance of objectivity; it maintains a full-time research correlation staff of specialists in highway transportation matters to bring the findings of research directly to those who are in a position to use them.

The program is developed on the basis of research needs identified by chief administrators of the highway departments and by committees of AASHO. Each year, specific areas of research needs to be included in the program are proposed to the Academy and the Board by the American Association of State Highway Officials. Research projects to fulfill these needs are defined by the Board, and qualified research agencies are selected from those that have submitted proposals. Administration and surveillance of research contracts are responsibilities of the Academy and its Highway Research Board.

The needs for highway research are many, and the National Cooperative Highway Research Program can make significant contributions to the solution of highway transportation problems of mutual concern to many responsible groups. The program, however, is intended to complement rather than to substitute for or duplicate other highway research programs.

This report is one of a series of reports issued from a continuing research program conducted under a three-way agreement entered into in June 1962 by and among the National Academy of Sciences-National Research Council, the American Association of State Highway Officials, and the U. S. Bureau of Public Roads. Individual fiscal agreements are executed annually by the Academy-Research Council, the Bureau of Public Roads, and participating state highway departments, members of the American Association of State Highway Officials.

This report was prepared by the contracting research agency. It has been reviewed by the appropriate Advisory Panel for clarity, documentation, and fulfillment of the contract. It has been accepted by the Highway Research Board and published in the interest of an effectual dissemination of findings and their application in the formulation of policies, procedures, and practices in the subject problem area.

NCHRP Project 3-2 FY'63
NAS-NRC Publication 1474
Library of Congress Catalog Card Number: 66-62659

FOREWORD

By Staff

Highway Research Board

This final report will be of special interest to traffic engineers and highway officials responsible for efficient operation of the existing urban street complex. It is an aggregation of current methods, equipment requirements, and costs for a computer-controlled traffic signal system. Instrumentation requirements for a real-time computer system for a small city are presented, and guidelines are developed to aid engineers in selecting control equipment. Control logic to be utilized by the computer in making signal change decisions is included. An annotated bibliography of previously completed research is contained in the appendix.

This investigation stems from the NCHRP project entitled "Surveillance Methods and Ways and Means of Communicating with Drivers." Only the findings of this project that pertain to traffic signal systems are reported herein. Studies involving freeway surveillance and the airborne observer in traffic control are presented in NCHRP Report No. 28.

As the traffic demand on the urban street system grows, the need for safe and efficient traffic signal timing continuously increases. In many urban areas severe traffic congestion has developed and new methods of alleviating the situation are desired. One approach that seems likely to ease the congestion problem is the utilization of a real-time digital-computer-controlled traffic signal system. Furthermore, due to the recent technological advancements in computer equipment, this type of traffic control, once considered exotic, seems likely to become in the near future a realistic alternative to be considered in the design and modernization of improved signal systems. For these reasons this research was conducted to apply state-of-the-art computer technology to traffic signal control, with emphasis being placed on applications for the small or medium-size city.

This research presents the most comprehensive analysis of control logic known to date coupled with a synthesis of the hardware required for a real-time closed-loop digital-computer-controlled traffic signal system. Although the system is based on a theoretical analysis, this report should extensively aid cities that are planning and developing digital-computer-controlled signal systems. It is highly desirable that the principles presented be applied to a real traffic system to validate this research.

The city of White Plains, N. Y., with a population of 50,000 and a trade area serving more than 250,000, was selected as a typical city to study. Utilizing this city as a model, a typical computer signal system is specified, equipment is selected, and costs are estimated for its 116 signalized intersections. Based on the analysis presented, it is anticipated that computer control requirements and costs may be estimated for other typical cities in this population class.

Control logic is derived based on the theories of minimum aggregate delay. The equations are presented in a readily usable form for solution by a computer system that is sensing traffic and making control decisions in real time. The report also provides needed information pertaining to acceptable criteria for the design and placement of traffic sensing devices.

In the control scheme presented, the signal change decisions are based on comparisons of time savings expressed in terms of relative delay and computer-derived predictions of future delays that will be incurred throughout the signalized network. In the event that street traffic is controlled utilizing the strategy presented, it is likely that major savings in time can be gained by the motorist traveling within the network of signalized intersections.

CONTENTS

FIGURES

TABLES

ACKNOWLEDGMENTS

Morton I. Weinberg, Head, Transportation Systems Section, had primary responsibility for the technical direction and administration of the project reported in both this report and NCHRP Report 28. Other Cornell Aeronautical Laboratory staff personnel working on the portions of the project reported herein were as follows:

Harvey Goldstein, Associate Electronics Engineer, who conducted and performed the detailed investigation into control techniques and philosophies applicable to the digital-computer-controlled traffic signal network, including the development of the control equations and statistical processes used therein, and assisted in determining some of the requirements for traffic surveillance equipment for this system. He was primarily responsible for Appendices E, G, H, and I, and had co-responsibility for Appendix F.

Terence J. McDade, Research Electronics Engineer, who performed the investigation into equipment necessary to bring traffic surveillance information into the digital computer and to implement the signal control commands. He was primarily responsible for Appendix C.

Robert H. Wahlen, Associate Electrical Engineer, who programmed portions of the control equations developed for the digital computer system as necessary to determine capacity and compute speed. He also investigated and selected typical digital computers that could be used for the network control system defined in Chapter Two, and had co-responsibility for Appendix F.

John P. Persico, Associate Mathematician, who conducted all data reduction runs and assisted in the model derivations and development.

Appreciation is expressed to the several individuals and companies who provided the project team with valuable assistance. Special acknowledgment is given to the following:

J. T. Hewton, Traffic Engineer, The Municipality of Metropolitan Toronto, for information on the equipment and progress of the computer-controlled traffic signal system in Toronto.

Roy A. Flynt, Commissioner of Planning, and Robert A. MacMonigle, (formerly) Traffic Engineer, both of the City of White Plains, N. Y., for traffic data with which to construct a traffic model.

Sherwood B. Bliss, Account Representative, International Business Machines Company, and Gary D. Haynes, Sales Engineer, Sperry-Rand Corporation, for information on their respective company's computer characteristics.

DIGITAL-COMPUTER-CONTROLLED TRAFFIC SIGNAL SYSTEM FOR A SMALL CITY

SUMMARY

This study is concerned with research on traffic control systems for an urban street complex. The present report describes the manner in which a control doctrine was developed and a digital-computer-controlled traffic signal system was synthesized for a small city. Equations were developed for assignment of right-of-way at an intersection that would result in minimum aggregate delay to motorists, within certain constraints. The equations are intended for solution in real time, using current information on traffic behavior within the street network. Traffic surveillance is to be accomplished by means of in-pavement loop detectors. Traffic and signal state information are transmitted to the central computer site via voice-grade telephone lines. A high-speed general purpose digital computer is to be used for processing the incoming data and "deciding" when commands to the signals to change phase should be generated. Changing of signal phase would be done by a pulse from the computer output acting on the local controller, which would be selected to act in the manual mode.

The evaluation of delay at an intersection is extended to include a subnetwork of not more than five intersections, the one being evaluated and the four downstream along the departure legs. This is typical of the intersection of two two-way streets.

The City of White Plains, N. Y., was selected as a traffic model. With a population of 50,000 acting as the core of a mercantile center serving 250,000, White Plains has 116 signalized intersections, for which it was estimated that 464 detectors would be needed to keep the computer adequately informed.

The minimum delay function was programmed in FORTRAN IV language, along with the subroutines for evaluation of quantities to be predicted, constants, and other inputs. In determining the storage capacity and computing speed requirements for the machine, all intersections were given an equal weight of importance. That is, it was assumed that a decision might be needed at each intersection at the end of each computing interval. The latter was assumed to be 2 sec, which is estimated to be adequate for real-time changes in the traffic complex. From the calculated computing machine requirements, two systems were selected from currently available equipment. The two had different purchase prices and maintenance charges, but over a 5½-year period either would have cost about $445,000, or $6,740 per month. It was estimated that programming and debugging the traffic control logic in the machine would cost about $25,000.

Detectors, tone generators, revisions to the local controllers, decoders, and input-output interface equipment at the computer were assembled into an inventory estimated to cost $229,000. Voice-grade telephone lines are expected to be used for communication of traffic data, signal states, and phase change commands. The rental fee for these lines was estimated to cost $2,000 per month.

Finally, it is suggested that the digital computer can be used for other purposes when the traffic load is light enough that the local controllers will give adequate service. At these times the computer will command the local controllers to take over so that it can be used as a business machine to serve the city's business and administration computing needs.

CHAPTER ONE

INTRODUCTION

The foreword of NCHRP Report No. 9 (*1*) closed with ". . . The next phase will include a study of freeway surveillance, a study of a digital-computer-controlled traffic signal network, further study of the requirements of an airborne-observer surveillance and control system, and a study of driver communication by aural and visual messages." Of the four tasks named, only the last has not been pursued. The results of the studies involving freeway surveillance and the airborne observer are presented in NCHRP Report No. 28 (*2*).

The current report deals with the digital-computer-controlled traffic signal network study. The approach that has been taken is to synthesize this type of signal system for a small city. A typical small city, White Plains, N. Y., furnished a fairly complete traffic model. A control doctrine of minimum delay was adopted and a generalized expression was developed to evaluate delay for a subnetwork of intersections. This, and the manner in which the expressions were implemented for real-time solution on a high-speed digital machine, is the subject matter of this report. In addition, an inventory of surveillance and communications equipment has been synthesized and the cost of the whole system, apart from the traffic signals and local controllers, has been estimated.

This research presents to the traffic engineer an idea of what and how many dollars, is involved in implementing a digital-computer control system for medium and small cities.

CHAPTER TWO

SYNTHESIS OF A DIGITAL–COMPUTER–CONTROLLED TRAFFIC SIGNAL SYSTEM

Traffic signals are generally employed at intersections as control devices, to alternately assign the right-of-way to orthogonally opposed traffic flows. However, determination of the apportionment of the total available time that is assigned to each flow or traffic movement will be a function of the type of control used and the predetermined criterion, or set of criteria, it is to implement. The selection of these criteria, or the figure-of-merit which is used to determine the operational effectiveness of the particular control, is in itself open to further commentary and research. That is, although most consideration has been given to the physical problem, many unknowns exist concerning the psychological stimuli to the human driver and their particular effect on his performance (e.g., accident proneness). In the study reported here the only criterion considered has been minimum aggregate delay to the vehicles, within the obvious safety constraints of limiting the amount of time a vehicle may be held by a red phase and the minimum time for a green phase, inasmuch as this approach currently appears most promising. However, the criterion of minimum psychological delay may deserve further consideration as more becomes known on the subject.

Because urban areas are in general characterized by their geometric grid network and associated heavy traffic volumes, they contain a high concentration of signalized intersections. It is to these intersections that the major portion of vehicular delay can be attributed. Without changing the physical characteristics of the network, the efficient operation of these signals then becomes a major factor in the reduction of the avoidable delay. The values of the control parameters selected (cycle length, splits, etc.) are at times critical if service to motorists is to be optimized. The term optimization in this context is used to mean the minimization of the motorists' travel time or delay.

By its very nature vehicular traffic is widely variable, with random fluctuations superimposed on regular hourly, daily, weekly and seasonal patterns. Therefore, the limitations of pretimed control of traffic signals are obvious. Hence, there have been numerous attempts to make the signal controls of both networks and isolated intersections sensitive to traffic fluctuations. As a consequence, there are many types of specialized traffic signal equipment and systems (traffic-actuated and traffic-adjusted) whose purpose is to expedite traffic through urban intersections. Although many such systems have produced good results, others have at times systematically aggravated rather than alleviated a specific traffic problem. In many cases, specialized systems implement discrete control philosophies which are artifices employed to obtain the minimal-delay results and are very limited in their ability to adjust to the rapid fluctuations of urban traffic. Although these special purpose systems sometimes are referred to as computer-controlled, the term "computer" is here reserved for the general purpose digital computer unless otherwise indicated.

To obtain a more desirable system having the sensitivity and flexibility to react to the uncertainty and variability of traffic flow that may occur within the time intervals of practical signal phase lengths, it is proposed that a general purpose digital computer directly implement the basic criterion of minimum aggregate delay. Another approach in the synthesis of such a system is to program the computer to emulate discrete traffic control techniques as currently employed by individual specialized systems (e.g., progressive, linked volume-density, etc.) With this method, as noted by Casciato and Cass (*3*), the computer would necessarily be capable of assigning various control subroutines to different intersections as dictated by traffic flow (sensors), time of day (clock), season (calendar), or the decisions of a human operator. Because it appears that implementation of the basic criterion is the more general approach and would yield the more responsive control (smaller reaction time to changes in traffic demands), synthesis of such a system is pursued in detail.

The control decisions of any such traffic-responsive system should result from continual sampling of the traffic flow characteristics and with such speed as to gather and process the pertinent data, calculate the decision, and actuate the controls in real time. To process the large quantities of data and perform the many calculations within the required time intervals, the system is synthesized employing a high-speed digital computer. Information describing the traffic flow and the signals displayed would be fed into the computer from strategically located detectors, whereas information concerning the street network geometrics (lane widths, number of lanes, location of detectors, etc.) would be stored within the computer for subsequent retrieval. The presence-type detectors would yield such information as to obtain the volume, occupancy, average speed, and average density across the lanes of coverage, in order to convey the state of the network to the computer. The computer, working in real time through the programmed equations implementing the minimum-delay criterion, would compute the optimum signal display at a given time for the entire network.

Hence, the proposed system may be viewed as a servomechanism (Fig. 1) in which the dynamic network of vehicles, pedestrians, weather, etc., is monitored with the limitations imposed by the detectors employed. The computer performs the calculations on the dynamic and static information and the resultant decisions or signal displays control the motorists and pedestrians.

The advantages and limitations of any such system, in vehicle-time of delay saved, are difficult to evaluate analytically. Specifically, it is suggested that the relative merits, when compared to less costly alternatives of some other system, be evaluated by computer simulation techniques. The ultimate test, of course, is in application of the system to the actual city being studied.

A detailed treatment of the material summarized in this chapter may be found in Appendices A through H. The equations given in this chapter are, for the most part, identified by the letter of the appendix and the serial number; thus, Eq. E-79 in the following text is Equation 79 of Appendix E.

DEFINITIONS AND NOMENCLATURE

The basic definitions used in the text of this report are assembled here for ready reference. Similarly, the various symbols used throughout the text are defined where they first appear, but are assembled here for the convenience of the reader. It should be noted, however, that inasmuch as most of the text equations and their discussions are summarized from the various appendices, and each appendix has its own nomenclature listing, the reader is advised to consult such listing for any symbols not contained in the following.

A precedent subscript (such as q, or A) indicates the intersection for which the term is applicable;

A following subscript (such as E, W, N, S) indicates the leg for which the term is applicable;

$\sim$ = an estimate of (the term to which it is applied);

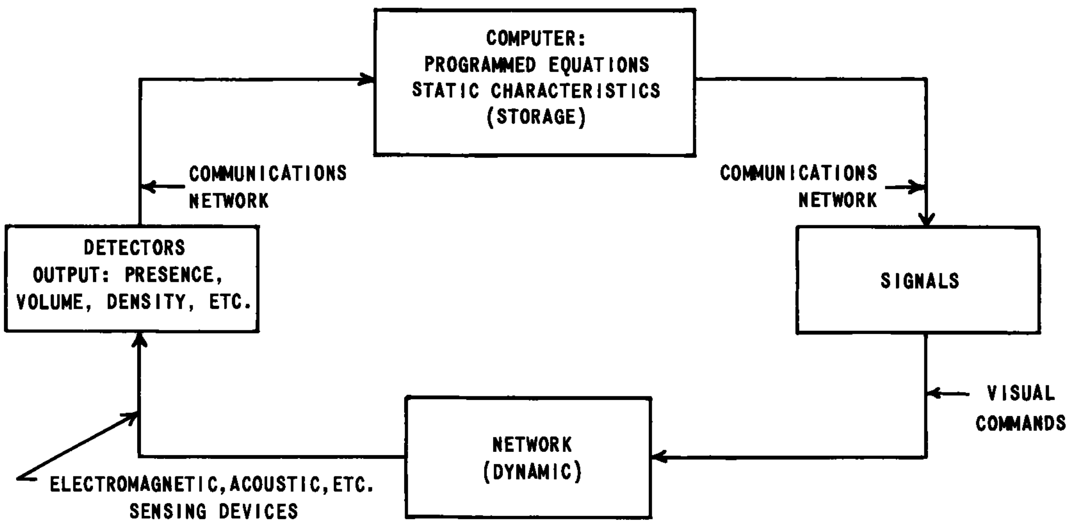


Figure 1. Adaptive-control feedback loop.

$A =$ designation of an isolated signalized intersection;
$_{A}[\ ] =$ with respect to intersection A;
$a =$ arrival rate, in vehicles per second;
$D =$ delay caused within the subnetwork;
$D_{acc} =$ delay attributed to the time necessary for vehicles stopped in queue to accelerate to free speed once they are released;
$D_{dec} =$ delay caused by deceleration of vehicles joining the end of the queue;
$D_{norm} =$ time required to traverse unimpeded the distance over which the delay equation applies;
$D_q =$ delay experienced by vehicles in queue waiting to be served by the intersection;
$D_T =$ relative delay suffered by traffic held on the non-moving leg;
$d =$ departure rate, in vehicles per second;
$d_{L^+} =$ departure rate, in vehicles per second, from a leg during the initial portion of the next phase, ϕ_G;
$i =$ a time interval, in seconds;
$j =$ a second time interval;
$K_1, K_2 =$ constants of proportionality;
$L = (\phi_R + \phi'_A + l)/\Delta t$, to the nearest whole number;
$l =$ lost time attributed to the initial acceleration period of vehicles;
$N =$ number of vehicles in queue;
$N_r =$ residual queue (number of vehicles) on a leg at the change of phase;
$n =$ an integral number;
q, r, s, t, u = signalized intersections in a subnetwork;
$S_D =$ delay saved due to departing vehicles during $n\Delta t$;
$S_q =$ savings accruing to vehicles forming in queue believed those released during $n\Delta t$;
$S_T =$ relative delay saved;
$T_1 =$ estimated delay saved by vehicles served (from two legs) during Δt, which otherwise would have to suffer amber, red, and lost time resulting from an immediate change in phase;
$T_2 =$ delay saved which accrues to vehicles served earlier in time as a result of previous departures;
$T_3 =$ additional relative saving in subnetwork, expressed as the sum of the individual savings at each of the neighboring signalized intersections;
$\Delta t =$ time interval for which the computer makes prediction;
$\Delta t_q =$ extended duration of the queue;
$\Theta =$ difference in relative delays and savings within a subnetwork for assumed alternatives at the intersection designated;
$\theta_{n\Delta t} =$ difference between relative delay saved by east-west traffic and relative delay suffered by north-south traffic during $n\Delta t$;
$\phi_A =$ total amber phase, in seconds;
$\phi'_A =$ portion of the amber phase that may be considered an extension of the red phase;
$\phi_G =$ green phase;
$\phi_{tG} =$ elapsed green phase time;
$\phi_R =$ red phase.

MATHEMATICAL MODEL

The starting point is observation of signalized intersection q (Fig. 2), whose departures are assumed to be "free" vehicles; that is, independent of any downstream intersection in terms of delay accrued. This situation may occur in practice when the nearest instrumented downstream intersection is far enough away to be considered independent. Further, it is assumed in the current example that only signals indicating stop (red), go (green), and clearance (amber) are available for display to diametrically opposed traffic streams (i.e., north-south or east-west symmetry). Hence, although exclusive pedestrian phases, leading phases, turning arrow-indicators, etc., are not considered, they are logical extensions of the subsequent development.

With the proposed control system the digital computer is periodically fed information from remote sensors that serve to sample the dynamic state of the intersection (i.e., current phase, approach volumes, departure rates, queue length, etc.) Based on the minimum delay criterion, the system uses all pertinent information, such as updated detector information, information from the computer's storage, and calculated predictions, to make a decision as to which signal display would be optimum.

As a result of these computer calculations, the local signals are controlled, by the transmission of the computer's decisions over the system's communication network to operate locally installed relay circuits. Existing local signal controllers will be maintained for use during periods of light traffic demand, in order that the computer may be free for other services, such as the city's business data processing needs. In addition, the local controllers may be viewed as a back-up system in case of an emergency situation, such as a computer breakdown or a transmission failure.

Practical Constraints

A traffic signal's phase length will be constrained by considerations other than the minimum-delay criterion. For instance, the minimum green time generally is chosen to provide pedestrian traffic with enough time to enter and cross the particular thoroughfare safely. If, indeed, this is the limiting factor for minimum green time, it becomes obvious that this value would vary with the intersection's dimensions and the number and type of pedestrians to be served. If it was possible to ascertain that there was no pedestrian traffic during a particular cycle, the computer would institute a minimum green time in accordance with the motorist's perception and reaction time and the vehicle's physical response time. The maximum red time is established so that no driver is required to wait an undue length of time before the signal changes in his favor. Hence, the signal is operated in accordance with the minimum delay criterion, within the aforementioned constraints.

Signal Displays Assumed

Among the first items to be considered are the types of calculations and information that would be necessary to come to a decision with respect to the state of the signal

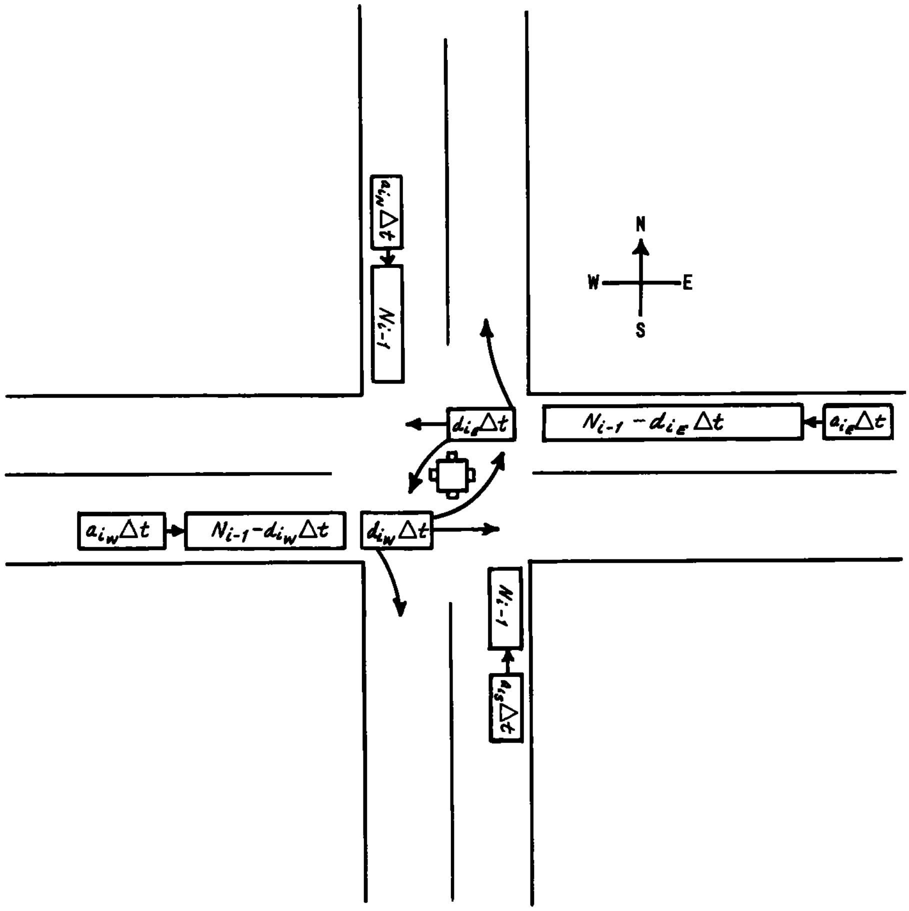


Figure 2. Events within a Δt *interval during the* $\phi_{G|E,W}$, $\phi_{R|N,S}$ *phase.*

displayed at an isolated intersection. The type and choice of action that may be taken (by changing the signal state) should first be considered. That is, from the assumed signal states there may be only green east-west or green north-south, while the opposite legs are red. The clearance (amber) interval is not considered as an option in the strict sense, because it is a required transitory state. However, the length of the amber interval for any one intersection may be made adjustable as a function of changes in road conditions, weather, and speeds.

Therefore, if the green signal as observed has an elapsed phase time greater than the minimum and less than the maximum, the alternative actions that may be considered are either to terminate the current signal state in favor of the other, or to extend the current phase for a specific interval of time. If this time interval is denoted by integral units of Δt, the control equations must provide for calculation of the relative aggregate delay which would be suffered to the vehicles on all approaches if the signals are changed immediately, or in Δt, or in integral multiples of Δt. For purposes of developing these equations it is assumed that the current phase is green to the east-west approaches, $\phi_{G|E,W}$, and the elapsed phase time, $\phi_{tG|E,W}$, is less than the maximum red phase, $\phi_{R|N,S_{max}}$, as displayed to the north-south traffic streams, and greater than the minimum green time, $\phi_{G|E,W_{min}}$.

Isolated Intersection

The computations performed for an isolated intersection yield the difference between two quantities—the relative delay saved by east-west traffic, $S_{T|E,W}$, and the relative delay suffered by the traffic being held on the north-south legs, $D_{T|N,S}$—as suggested by Miller (*4*). The two quantities result from an extension of the current east-west green phase by some increment of time, $n\Delta t$, as compared to immediate termination of the phase. This difference, $\theta_{n\Delta t}$, is defined as

$$\theta_{n\Delta t} = {}^{n\Delta t}\,[S_{T|E,W} - D_{T|N,S}] \tag{1}$$

and is the objective function which should be made negative before a change in signal phase is warranted.

If the value of $\theta_{\Delta t}$ is negative, and if no successive time increments are investigated, it is known* only that the net

* Although the term "known" is used here, it should be noted that the quantity θ is a prediction of what might occur some time in the future and as such is subject to the errors inherent in the prediction model. For example, it is not known absolutely how many vehicles will depart or arrive during the next $n\Delta t$ seconds, but a good prediction of these values may be obtained easily from past and current observations. These predictions are denoted in the text by the tilde; for example, $\tilde{d}$ is a predicted departure rate.

delay incurred would be greater by extending the current phase by Δt than by immediate termination. It is necessary to evaluate aggregate delay for proposed extensions of $2\Delta t$, $3\Delta t$, . . ., $n\Delta t$, to determine if these alternative actions would yield less delay than changing the signal state immediately.

Isolated Savings Expression

In order to evaluate the savings expression, $S_{T|E,W}$, it is considered as the sum of two separate terms. The first term, S_D, is a result of the delay saved due to the departing vehicles during the proposed extension, $n\Delta t$. The second term, S_q, is the resultant savings that accrue to the vehicles forming in queue behind those that have been released from the stop-line of the intersection during the proposed extension. (The subsequent formation of the queue is served at an earlier time once the signals have again changed in their favor.) The relative delay saved for intersection q may now be written as follows:

$$ {}_qS_{T|E,W} = {}_q\{S_D + S_q\}_{E,W} \qquad \text{(D-1)} $$

Expressing S_D by multiplying the total number of vehicles predicted to depart the east and west legs during the proposed extension, $n\Delta t$, by the prediction of the total lost time gives

$$ S_D = \left\{ \sum_{i=1}^{n} (\tilde{d}_{iE}\Delta t_i + \tilde{d}_{iW}\Delta t_i) \right\} \{\tilde{\phi}_{R|E,W} + \tilde{\phi}'_{A|E,W} + \tilde{l}_{E,W}\} \qquad \text{(G-40)} $$

The second term of Eq. D-1 expresses in vehicle-seconds the relative delay saved which accrues to those vehicles served earlier in time as a result of the previous departures. The factor must include computation of the total number of arrivals on both approach legs during the next phase, $\phi_{R|E,W}$, plus those remaining in queue, and then multiplication of this sum by the time that these vehicles are closer (in time) to service. Hence,

$$ S_{qE,W} = \left\{ \left(\sum_{i=1}^{L} \tilde{a}_{iE}\Delta t_i \right) + N_{,E} \right\} \frac{\sum_{i=1}^{n} \tilde{d}_{iE}\,\Delta t}{\tilde{d}_{L^+E}} + $$

$$ \left\{ \left(\sum_{i=1}^{L} \tilde{a}_{iW}\,\Delta t_i \right) N_{,W} \right\} \frac{\sum_{i=1}^{n} \tilde{d}_{iW}\,\Delta t_i}{\tilde{d}_{L^+W}} \qquad \text{(G-42, G-43)} $$

in which the various a_i and d_i terms are defined by Figure 2.

Isolated Delay Expression

The delay, $D_{T|N,S}$, caused to the vehicles on the north and south legs, respectively, by deferring the start of the $\phi_{G|N,S}$ phase for a period of $n\Delta t$ may be represented by three separate delay terms. They are, D_{dec}, caused by the deceleration of vehicles to join the end of the queue, D_q, experienced in the queue waiting to be served by the intersection, and D_{acc}, attributed to the time necessary for those vehicles stopped in queue to accelerate to free speed once released. However, because the term delay was taken to mean the additional travel time above that normally incurred, another factor, D_{norm}, representing the average number of vehicle-seconds required to traverse the distance over which the delay equation applies, must be included. The combined expression for delay for the north and south legs is

$$ D_{T|N,S} = D_{dec|N,S} + D_{q|N,S} + D_{acc|N,S} - D_{norm|N,S} \qquad \text{(D-9)} $$

Figure 3 expresses the queue length for a particular approach leg (number of vehicles in queue) as a function of time.

Although all vehicles required to stop at an intersection will suffer some delay, when compared to a free-flow condition, only those (additional) vehicles impeded by the Δt_q extended duration of the queue are considered in the relative delay calculations of the D_{dec} and D_{acc} terms. If it is assumed that all vehicles have similar deceleration and acceleration characteristics (*5*), both the aforementioned delay terms will be proportional to the number of additional vehicles which must fall in queue. With constants of proportionality of K_1 and K_2 assumed for $D_{dec\ S}$ and $D_{acc\ S}$, respectively,

$$ D_{dec\ S} = K_1 \sum_{i=k_S}^{k'_S} \tilde{a}_{iS}\,\Delta t_i \qquad \text{(D-14)} $$

$$ D_{acc\ S} = K_2 \sum_{i=k_S}^{k'_S} \tilde{a}_{iS}\,\Delta t_i \qquad \text{(D-15)} $$

Next is examined the term D_q, expressing the difference in vehicle-seconds suffered in the queue for the two alternatives. The queue size, N, in general may be expressed as a continuous function of time $N(t)$. This representation is physically realistic, because vehicles do not jump into or out of queue but rather flow (e.g., ½ veh/sec). Hence, Figure 3 may indeed express the general queue size on an approach leg to an intersection for two specific cases of the alternatives considered.

The term D_q may be expressed mathematically as the difference of two integrals; thus,

$$ D_q = \int_0^{k'\Delta t} N'(t)\,dt - \int_0^{k\Delta t} N(t)\,dt \qquad \text{(D-16)} $$

However, in order to apply this equation it is assumed that the departure and arrival rates over an interval, Δt, are constant. Therefore, the exact discrete representation of D_q is written as

$$ D_q = \sum_{j'=1}^{k'_S} \left\{ \left[N_{I_j} + \sum_{i'=1}^{j'} (\tilde{a}_{i'} - \tilde{d}_{i'})\,\Delta t_{i'} - (\tilde{a}_{j'} - \tilde{d}_{j'}) \frac{\Delta t_{j'}}{2} \right] \Delta t_{j'} \right\} $$

$$ - \sum_{j=1}^{k_S} \left\{ \left[N_{I_j} + \sum_{i=1}^{j} (\tilde{a}_i - \tilde{d}_i)\,\Delta t_i - (\tilde{a}_j - \tilde{d}_j) \frac{\Delta t_j}{2} \right] \Delta t_j \right\} \qquad \text{(D-18)} $$

The previously mentioned equations, particularly Eq. 1, may be implemented in real time if the traffic quantities of predicted arrivals, departures, phase lengths, and measured queue length, etc., are obtained within such time as to permit computation of a decision within Δt seconds.

Network Considerations

The isolated equations for θ_n are applicable in all cases where the delay suffered at the next signalized intersection

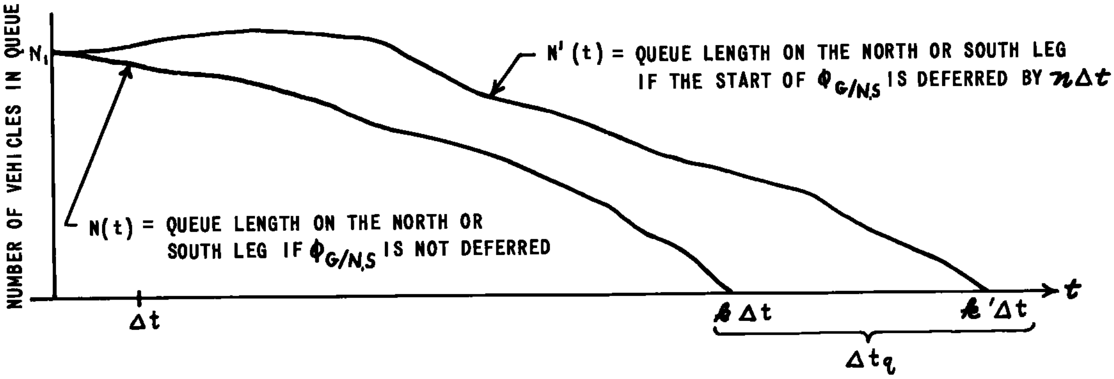


Figure 3. Queue length vs time.

encountered enroute by the vehicles released from q may be neglected in considering the state of q. Although this is used as the definition of an isolated intersection, it should be noted that arrivals to intersection q may be dependent on a neighboring upstream signalized intersection (see Fig. 4). In this case the predictions of the arrivals to q will reflect such influences.

The choice to neglect a certain portion of the surrounding network with respect to the decision to be made at intersection q (Fig. 4) is the result of practical considerations and, as was shown, does not imply that such an intersection is literally isolated. (The intersection may be influenced by conditions at other intersections and in turn, may itself influence conditions at surrounding intersections.) It does, however, indicate that the intersection is being treated as isolated because departing vehicles may have to travel relatively large distances before reaching another signalized intersection or measurement station, or must pass through a number of unsignalized intersections before reaching the next signalized intersection. "Isolation" may be caused by travel over a relatively short distance where there may be a significant amount of friction, such as sources and/or sinks, or any combination of various degrees of the aforementioned conditions. It should be noted that in all these instances it may be possible to predict with a high degree of certainty that the vehicles will arrive at a particular point; but at what time they will arrive has associated with it a very large uncertainty. In such instances it is apparently futile to link the intersection being evaluated with the neighboring ones by estimating the delay incurred at the signalized intersection downstream from the released vehicles. However, in urban areas, and in particular within the central business districts, such isolated intersections are rarely found. This is attributed to the geometry of the grid network and to the vehicular and pedestrian volumes attracted to the area, necessitating a high density of signalized intersections, and the more or less ordered flow of traffic which is prescribed by the enforcement of various traffic ordinances.

A question which logically arises at this point is: "How many signalized intersections downstream from the vehicles released from the intersection being evaluated should be considered in the network delay equation?" In the most general case of an intersection of two two-way streets each vehicle has the choice of three paths (straight through, left, or right turns). As the number of intersections being considered is increased, the probability of the released vehicles reaching the farthest intersection downstream (in the original direction of travel) is decreased as a function of the distributions of the turning movements at each intersection. Further, one may examine the situation whereby turning movements are prohibited and the arrivals at a

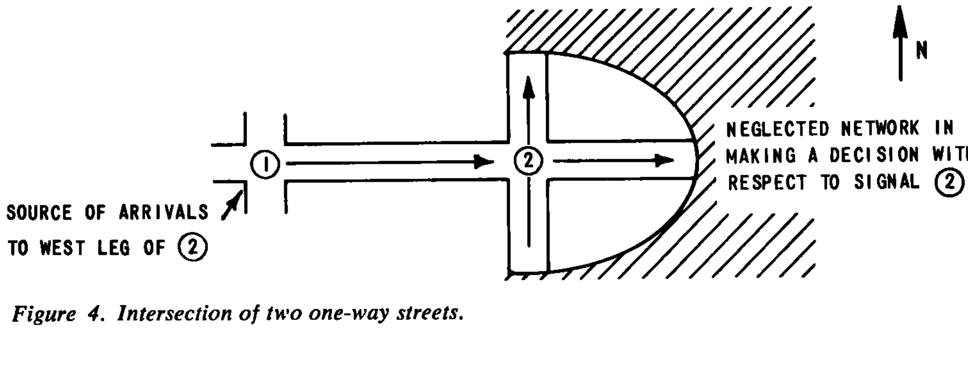


Figure 4. Intersection of two one-way streets.

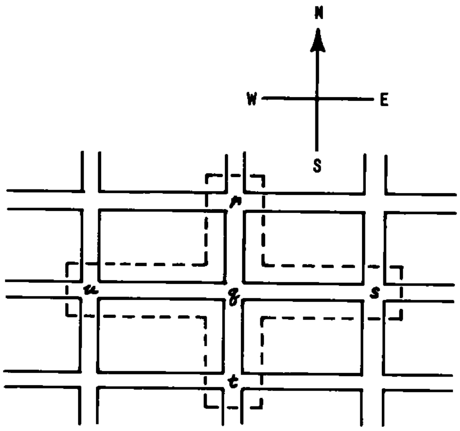


Figure 5. Subnetwork (q, r, s, t, u) of signalized intersections.

point downstream are certain but are dispersed only by the variance on the mean velocity which has associated with it the variance of the arrival times, accounting for the distance traveled. With the inclusion of one signal, or even one stop sign, controlling another intersection between the intersection being evaluated and the one downstream, the standard deviation of the predicted arrival times may be greater than the duration of the green phase (for the downstream signal), thus introducing a significant error in the predicted delay.

For practical reasons, such as those previously described, network equations are generally limited to the immediate neighboring signalized intersections, normally consisting of four pairs of intersections, each including the intersection being evaluated.

Subnetwork Problem Statement

This section develops a set of general equations that will measure the relative delay to vehicles, currently being controlled at intersection q (Fig. 5), that includes the predicted delay when these vehicles reach the downstream intersections, r, s, t, and u, for the control alternatives assumed to be available at q. Although the minimum-delay solution is sought in terms of the five intersections of two two-way streets, it is realized that many other network configurations are possible. However, the most common ones will simply be a modification of the subnetwork shown. Peculiarities to the sample city are discussed later.

The problem is to decide when is the optimum time to change the signal at q, assuming a symmetrical signal display, where "optimum" signifies the minimum aggregate delay accrued within the subnetwork to those vehicles being controlled at intersection q.

The current conditions at q and the state of the rest of the network (phase, phase time elapsed, queue lengths, approach rates, departure rates, various predictions based on past measurements, etc.) are used to implement the minimum-delay criterion. The isolated case was concerned solely with the delay associated with intersection q. Here, the equations are extended by considering the delays incurred in the subnetwork contained within the dashed lines.

It is assumed that the signal at q is green to east and westbound traffic, phase ${}_{q}\phi_{G|E,W}$. It must be decided whether the current phase should be extended by $n\Delta t$ or terminated immediately. To evaluate the alternatives, the delay accrued by the vehicles up to the time that they depart from the subnetwork (at the perimeter of the dashed enclosure) must be considered. The implementation may take the form of evaluating either the difference in times of service at the perimeter (exit of subnetwork) or the sum of delays incurred at each intersection along the route. The latter approach is pursued here. Hence, the total relative delay or savings is considered to be the sum of contributions from, at most, two signalized intersections for any particular vehicle. Therefore, the $\theta_{n\Delta t}$ term for the isolated intersection will be applicable to the total development and will be the starting point.

Savings Within Subnetwork

The $S_{T|E,W}$ term is only part of the total savings within the subnetwork, as this is the savings of delay at intersection q only. The additional relative savings within the entire subnetwork will be expressed as the sum of the individual savings at each of the neighboring signalized intersections, as follows:

$$ {}^{n\Delta t}{}_{q}S' = {}^{n\Delta t}\left({}_{q,r}S + {}_{q,s}S + {}_{q,t}S + {}_{q,u}S\right) \qquad \text{(E-1)} $$

It should be noted that the pre-subscript q refers to the situation whereby the savings equation compares the relative benefits of the alternative decisions to be made at intersection q. The double pre-subscript refers to the intersection within the subnetwork where the saving occurs. Further, the ${}_{q}S'$ term does not express the total relative savings* within the subnetwork q, r, s, t, u. This term is denoted by

$$ {}^{n\Delta t}{}_{q}S = {}^{n\Delta t}\left({}_{q}S' + {}_{q}S_{T|E,W}\right) \qquad \text{(E-2)} $$

where the last term is defined by Eq. D-1. Each of the contributions to ${}^{n\Delta t}{}_{q}S$ is now expressed in terms of the difference in delays suffered at the neighboring intersections which would occur if the vehicles are released as a result of the extension of the current green phase at q, or after being held for the duration of the next red phase, ${}_{q}\tilde{\phi}_{R|E,W}$. Therefore,

$$ {}^{n\Delta t}\left[{}_{q,r}S = {}_{q,r}D^{\tilde{\phi}_{R|E,W}} - {}_{q,1}D^{o}\right] \qquad \text{(E-3)} $$

where ${}_{q,r}D^{\tilde{\phi}_{R|E,W}}$ is the delay predicted to be incurred at r, to the vehicles released from intersection q after waiting for the next red phase, ${}_{q}\phi_{R|E,W}$, and choosing their path through r; and ${}_{q,r}D^{o}$ is the delay predicted to be incurred at r, to the same vehicles that are released during the current proposed extension and also desire service at r. The pre-superscript $n\Delta t$ indicates that the evaluation is performed for a particular value of n. Similar equations may be written for the relative saving at s, t, and u.

* This value may be either positive or negative. If it is negative, then more delay is predicted to be incurred by the vehicles from the east-west legs of q at the neighboring intersections r, s, t, u, if released during the proposed extension, $n\Delta t$, than if held at q for the subsequent red phase, ${}_{q}\phi_{R|E,W}$, and then released. The term "savings" is always applied to the vehicles on the legs currently receiving the green signal.

Delay Within Subnetwork

Up to this point the "savings" of delay, ${}_{q}S$, within the subnetwork for those same vehicles predicted to receive service at intersection q (east-west legs), during the alternatives available at $q(0,\tilde{\phi}_{R|E,W})$, have been considered by comparing the delay experienced downstream at the neighboring intersections.

The expression for the relative delay*, ${}_{q}D$, caused within the subnetwork is defined in terms of the previously defined delay at q, $D_{T|N,S}$, and the relative delay accrued to the vehicles at the neighboring intersections which are predicted to be released from the north and south legs of q.

Again, the predicted consequences of the two alternative actions at intersection q—(1) Terminate current phase, initiate ${}_{q}\phi_{G|N,S}$ now, or (2) Defer action by the extension of ${}_{q}\phi_{G|E,W}$ through $n\Delta t$—are compared. First, the delay caused within the subnetwork is expressed as

$$^{n\Delta t}{}_{q}D = {}^{n\Delta t}{}_{q}D' + {}_{q}D_{T|N,S} \qquad \text{(E-44)}$$

where

$$^{n\Delta t}{}_{q}D' = {}^{n\Delta t}({}_{q,r}D + {}_{q,s}D + {}_{q,t}D + {}_{q,u}D) \qquad \text{(E-46)}$$

From a previous definition, the first pre-subscript indicates the intersection to which the decision is applied and the second indicates the particular intersection at which the relative delay is considered.

Now, a particular intersection, r, within the subnetwork q, r, s, t, u, is investigated. By referring to Eq. E-10, the relative delay at r may be expressed as

$${}_{q,r}D = {}_{q,r}(D^{n\Delta t} - D^{\phi_{A|E,W}})_{S|N,S} \qquad \text{(E-47)}$$

where

${}_{q,r}D_{S|N,S}{}^{n\Delta t}$ = delay suffered on south leg of r, to those vehicles released from the north-south legs of intersection q, if the current phase ${}_{q}\phi_{G|E,W}$ is extended by $n\Delta t$; and

${}_{q,r}D_{S|N,S}{}^{\phi_{A|E,W}}$ = delay suffered on south leg of r, to the vehicles released from the north-south legs of intersection q, if the current phase ${}_{q}\phi_{G|E,W}$ is terminated immediately.

The D^{ϕ_A} term, which is similar in form to $D^{n\Delta t}$, $D^{\tilde{\phi}_{A|E,W}}$, and D^{o} of Eq. E-3, is presented as the sum of its component delay terms:

$$D^{\phi_A} = \{D_{dec} + D_q + D_{acc} - D_{norm}\}^{\phi_A} \qquad \text{(E-8)}$$

Subnetwork Objective Function

The basis for choosing a particular action from the alternatives available is considered. A function, ${}_{q}\Theta_{n\Delta t}$, may be developed which expresses the difference in relative delays and savings within a subnetwork for the two assumed alternatives at intersection q. In terms of a previously defined notation, the following equation is introduced:

$${}_{q}\Theta_{n\Delta t} = {}_{q}\theta_{n\Delta t} + {}^{n\Delta t}{}_{q}S' - {}^{n\Delta t}{}_{q}D' \qquad \text{(E-79)}$$

or

$${}_{q}\Theta_{n\Delta t} = {}^{n\Delta t}_{q}\{({}_{q}S - {}_{q}D) + ({}_{r}S - {}_{r}D) + \ldots ({}_{u}S - {}_{u}D)\} \qquad \text{(E-80)}$$

The computer will perform the following test:
If

$${}_{q}\Theta_{\Delta t} \geq 0 \qquad \text{(E-81)}$$

extend the current phase, because more or equal delay is predicted to be accrued by terminating the current phase. However, if

$${}_{q}\Theta_{\Delta t} < 0 \qquad \text{(E-82)}$$

less delay would be incurred by changing the phase now than Δt from now. However, it may be necessary to calculate ${}_{q}\Theta_{2\Delta t}$, ${}_{q}\Theta_{3\Delta t}$, . . ., ${}_{q}\Theta_{n\Delta t}$, because less delay may be incurred at these other times. It should be noted that if any of the subsequent ${}_{q}\Theta$'s are positive the procedure may terminate and the particular extension whose Θ yields a positive quantity is accepted.

A descriptive flow diagram of the calculations to be performed at each intersection once every Δt seconds is shown in Figure 6; its mathematical equivalent, in Figure 7.

MEASURED TRAFFIC QUANTITIES

To implement the developed minimum-delay equations, the computer must "know" both static information and the value of the current dynamic traffic parameters. The static information is relatively fixed and may be entered into the computer and updated infrequently (weekly, monthly, annually). It represents physical quantities that need not, or in some cases cannot, be included in the real-time feedback loop. Included in the table of static information would be the number and definition of the intersections contained within each subnetwork, street widths, number of lanes, one-way streets, constraints (minimum green, maximum red), number of intersections within network under computer control, street lengths, etc.

The dynamic information, on the other hand, may fluctuate widely between cycles, minutes, or even seconds. Hence, this information must be presented to the computer within such time as to update most of these measurements before the next decision is made. Not to do this would be to base decisions on outdated information, which would produce less than the optimal solution to the current traffic demands and could lead to the undesirable result of maximizing the delay. This situation appeared to have happened with some special-purpose systems which were either not sensitive enough to traffic flows or were "maladjusted" in some other way.

It is proposed that, for the general intersections, approach and departure rates on each leg would be presented to the computer once every Δt seconds. Also, it should be noted that the information obtained from the detectors does not actuate any of the signals directly (Fig. 8). Instead, all information is fed to the computer, which in turn performs the calculations necessary to evaluate the

* Although this term is denoted as "relative delay," it should be noted that its sign may be positive or negative. If it is negative, vehicle-seconds are predicted to be saved within the subnetwork by the north-south departures if deferred an additional $n\Delta t$ seconds.

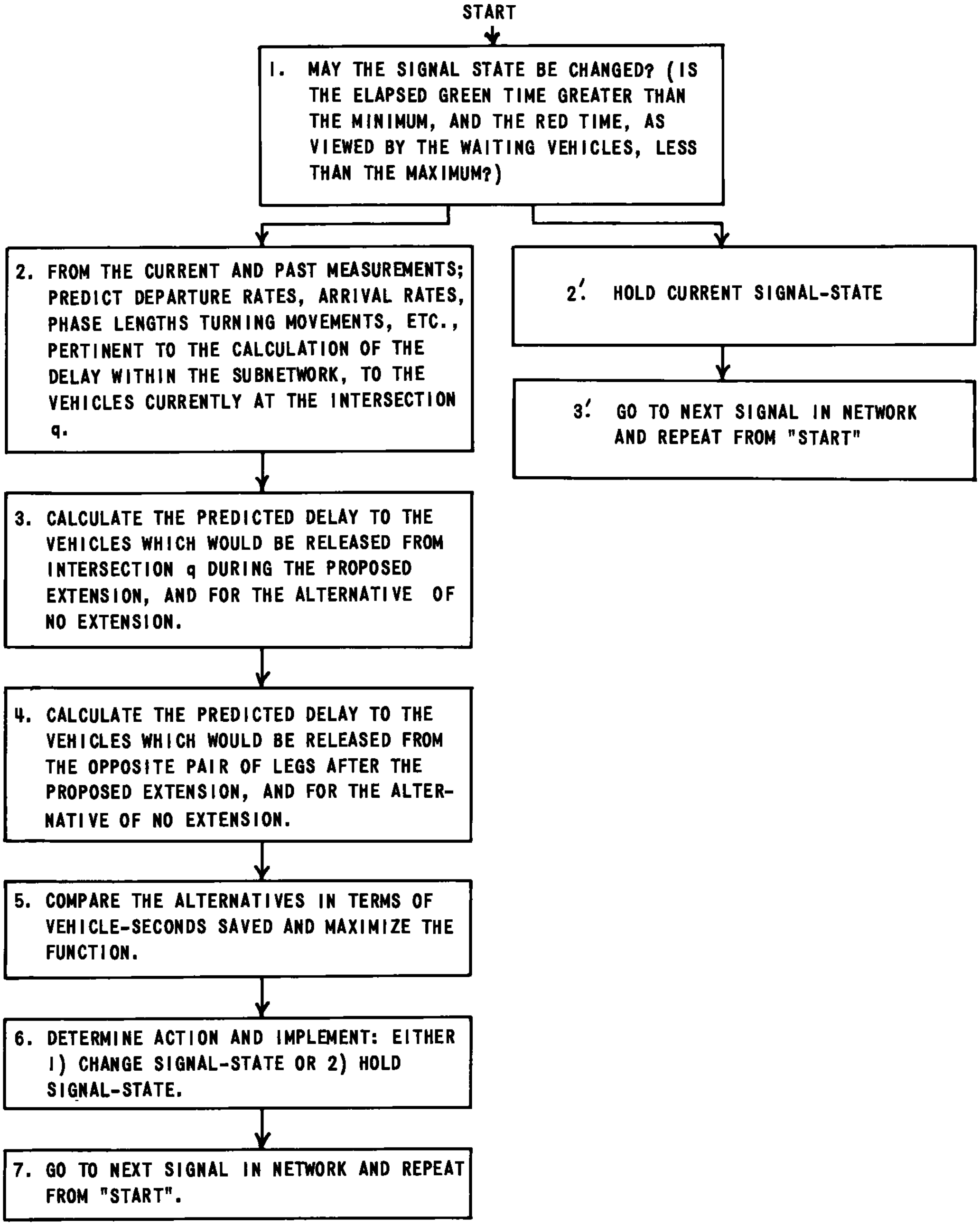


Figure 6. Descriptive flow chart of the implemented criterion.

entire network. Information concerning the current signal display also is included. Calculations are performed to obtain the current queue lengths and elapsed phase time every Δt seconds. Velocity calculations also may be included with the same basic detector information, if desired.

PREDICTIONS

Predictions of the necessary quantities (future departure and arrival rates, cycle and phase lengths, etc.) are updated as a result of each new measurement. Models may be selected for various portions of the day to anticipate any regular fluctuations.

For example, predictable fluctuations may include the onset of rush-hour periods, sporting events, holiday shopping, parades, etc., information which may be entered into the computer at the central site. This does not prescribe a particular progression, or offsets, splits, etc., but allows the system to remain in an adaptive control mode. However, the prediction techniques are thereby altered to be more sensitive to the forthcoming situation, which is assumed to be known *a priori*.

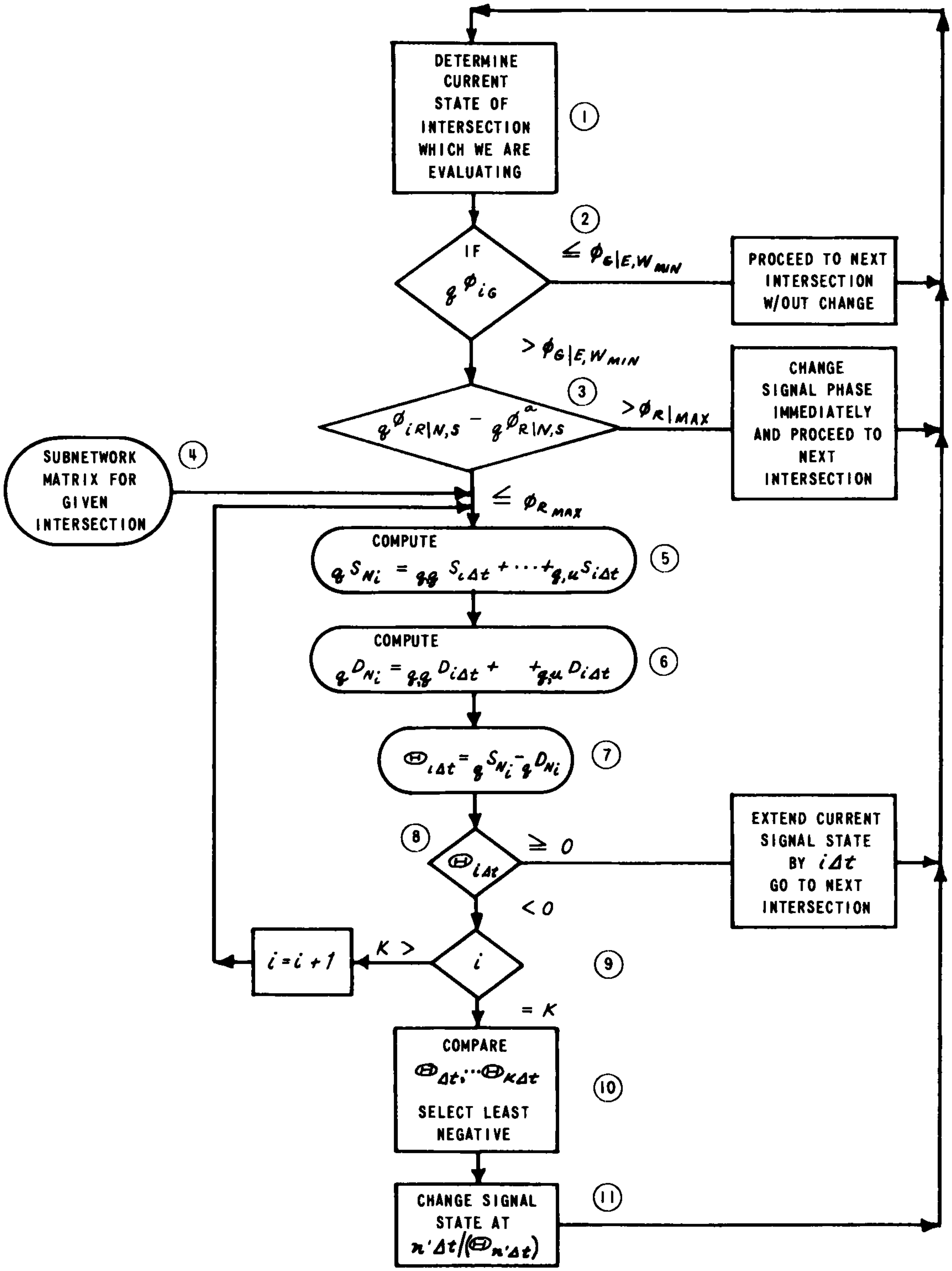


Figure 7. General flow chart of saving and delay equation criterion.

These predictions and the models employed determine the sensitivity of the network's signal response to the fluctuating traffic demands. Hence, the merit of the proposed (implemented) control system obviously relies on judicious selection of the prediction techniques.

One of the problems that plagues even the most advanced traffic signal system known to be operational is its reliance on certain expected values. These averages, which are used as predictions of important traffic quantities, may have the added flexibility of being varied as a function of time (clock), season (calendar), observations (human), etc. However, these predictions are generally based solely

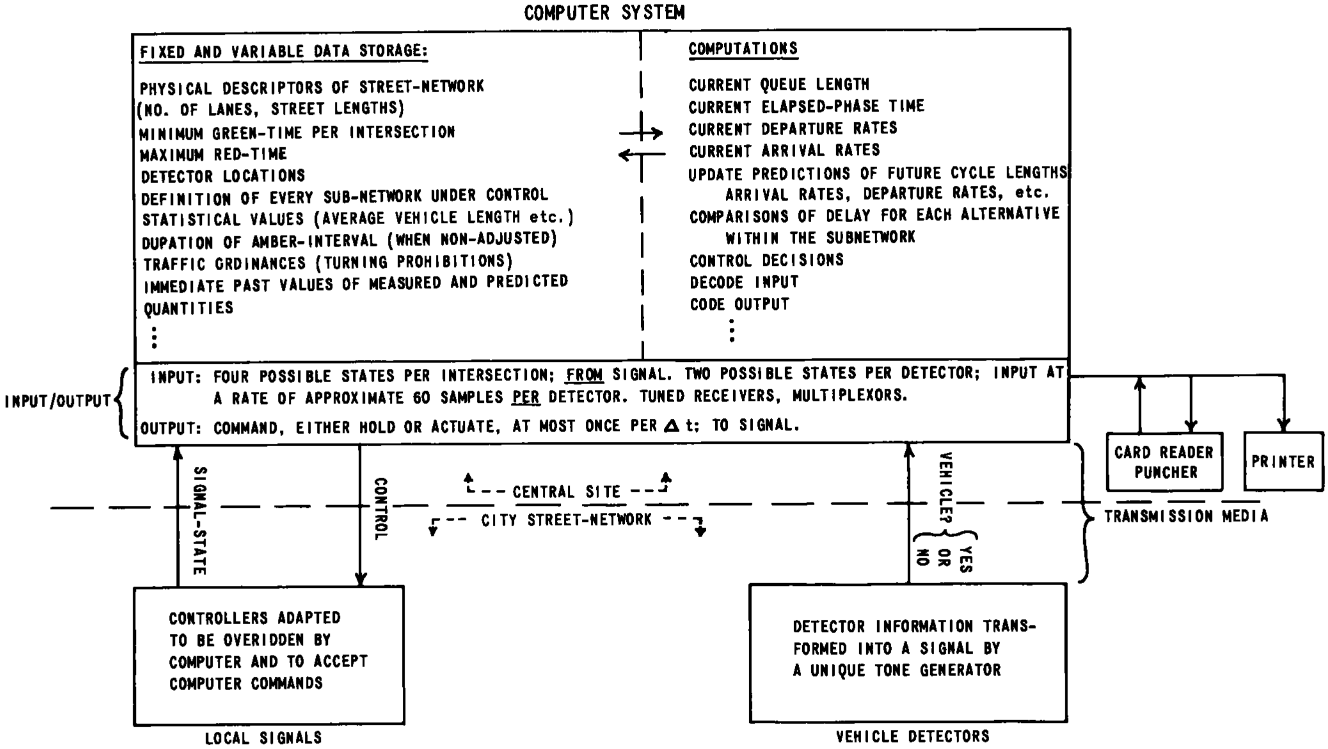


Figure 8. Integrated control system, both central site and street network.

on past measurements observed during the last hour, day, or year. Furthermore, the measurements may have been taken at other than the particular intersection or locale to which they are applied. Such a philosophy might be acceptable for a freeway or highway where the statistics at various points may be assumed invariant as a function of location. Applying this approach to urban traffic predictands could lead to poor control.

It is the very nature of the traffic characteristics experienced within urban areas to include large variations in the important quantities necessary for effective real-time control, which cannot be forecast with outdated observations. These wide fluctuations occur not only from day to day, or hour to hour, but also within seconds. For example at the stop line at the north leg of a hypothetical intersection departures have been flowing at the saturation rate since the end of the starting transient. Suddenly a vehicle desires to turn left, but must wait for a gap in the conflicting traffic. In a matter of seconds the flow could have been reduced to zero. One problem suggested by this example is that of obtaining a good prediction of the departure flow subsequent to the turning movement. Such a prediction would best be made by including all quantities known to be correlated with the flow rate and the most recent measurements attained by the system. It should be noted that not all of the necessary predictions to be made are as critical or as fluctuating as in the example described.

The variances on the traffic quantities are believed to be attributed to both attainable and unattainable (due to instrumentation limitations) complex factors. For example, departure rates at a particular leg of an intersection may be affected by time of day, day of week, season, weather, turning movements, actual delay experienced in the network, length of queues, special local events, international news, personal and emotional experiences, etc., some of which may be correlated with each other. Based on the foregoing discussion and the traffic data analyzed, it is believed that the proposed system should lend itself to updating techniques that include the most recent information, when necessary. Hence, as the computer continues to interrogate the remote instrumentaton (sample the dynamic network), the predictions would be revised to reflect these additional data. In this manner, an attempt will be made to make the system respond to both sudden changes in conditions affecting the predictions and the more subtle variations which occur throughout the day.

Although it is not the purpose of the current study to empirically validate various prediction models, such a study would be necessary for the successful implementation of a control system, and should be performed (in part) before a system is made operational. In view of the foregoing limitation, the authors investigated a few theoretical prediction methods that appeared to be reasonable in their applicability to the developed equations, and also relied on the empirical validation of other research work in this area, such as reported by Miller (*4*) and Greenshields and Weida (*12*).

In selecting the prediction models the computational requirements of the digital computer were considered, because the prediction is most valuable if it can be computed simply and quickly so as to be applied within the real-time limitations.

It is believed that good predictions of the necessary traffic quantities (those with small possible errors) may be obtained by combining simple linear regression, multiple linear regression, and average values, with a weighted moving average technique. The complexity of the particular model proposed for each of the quantities to be predicted will depend in part on the apparent physical reasonableness of the assumptions, the prediction interval required (confidence level), and the instrumentation as proposed.

Use of Averages as Predictions

There are various types of averages, which at first inspection appear to be good for predictions in general. Specifically, it is desired to predict certain traffic quantities within the urban network that may be implemented by a digital computer in making real-time control decisions. With respect to this application, the properties of these averages (and predictions in general) appear to be as presented in the following. The averaging techniques should:

1. Include as many of the past observations as practical.
2. Weight the observations to give more importance to the more recent rather than to the earlier observations.
3. Require simple calculations in order to minimize computer size and solution time (real-time applicability).
4. Be able to include the most recent observations without complexity (easily updatable).

In view of the foregoing, an exponentially weighted moving average is proposed for use in predicting various traffic quantities. These averages have been widely used for the operation of inventory control systems. Exponentially weighted moving averages are used to extrapolate a sales time-series (forecast sales), because they can be made quickly, easily, adaptable to electronic computers, with a minimum amount of information, and to introduce current sales information (*11*). Furthermore, Miller (*4*) presents some empirical studies that to some degree validate the application of such averages to the predictions of certain traffic quantities needed for urban signal control.

In order to update the exponentially weighted moving average to reflect the additional information of the most recent observation, y_0, one must simply add to this measurement the product of the weight, ρ, and the difference between the measurement and its previous prediction, $(\tilde{y}_1 - y_0)$. Figure 9 indicates the time sequence of events occurring when the current observation is to be included. This technique also lends itself to simple and multiple linear regression, where current measurements of the predictor quantities must be included.

Proposed Prediction Models

The following is a discussion of the manner in which the values of two predicted quantities may be obtained. For one of these quantities an alternate scheme is shown.

The first prediction is that of "lost" time to a particular

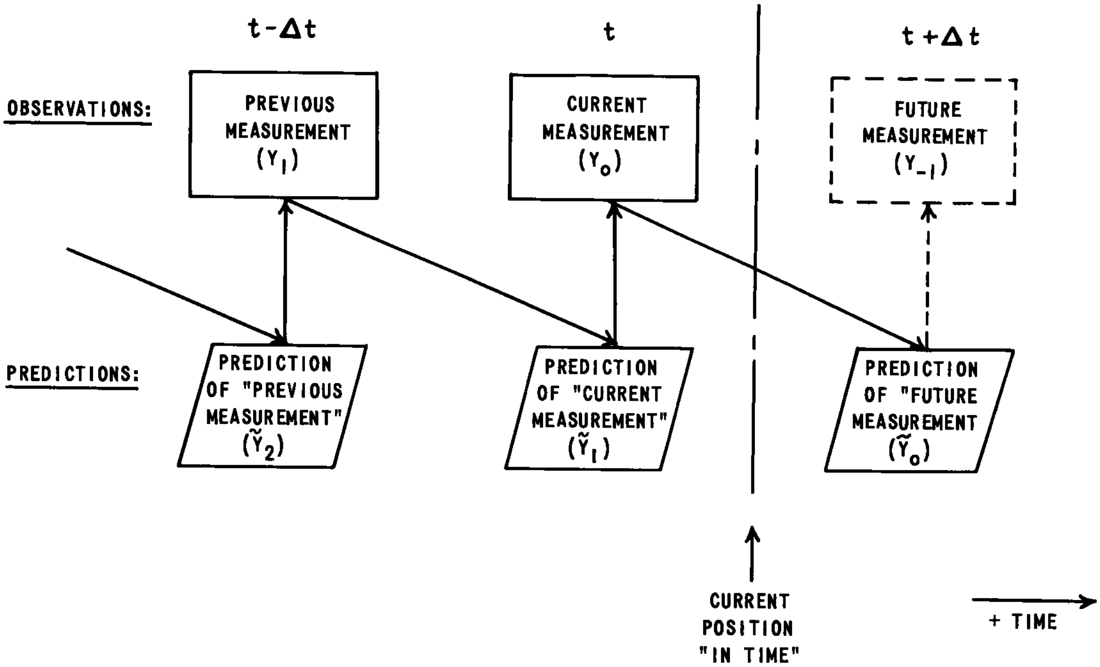


Figure 9. Updating prediction technique.

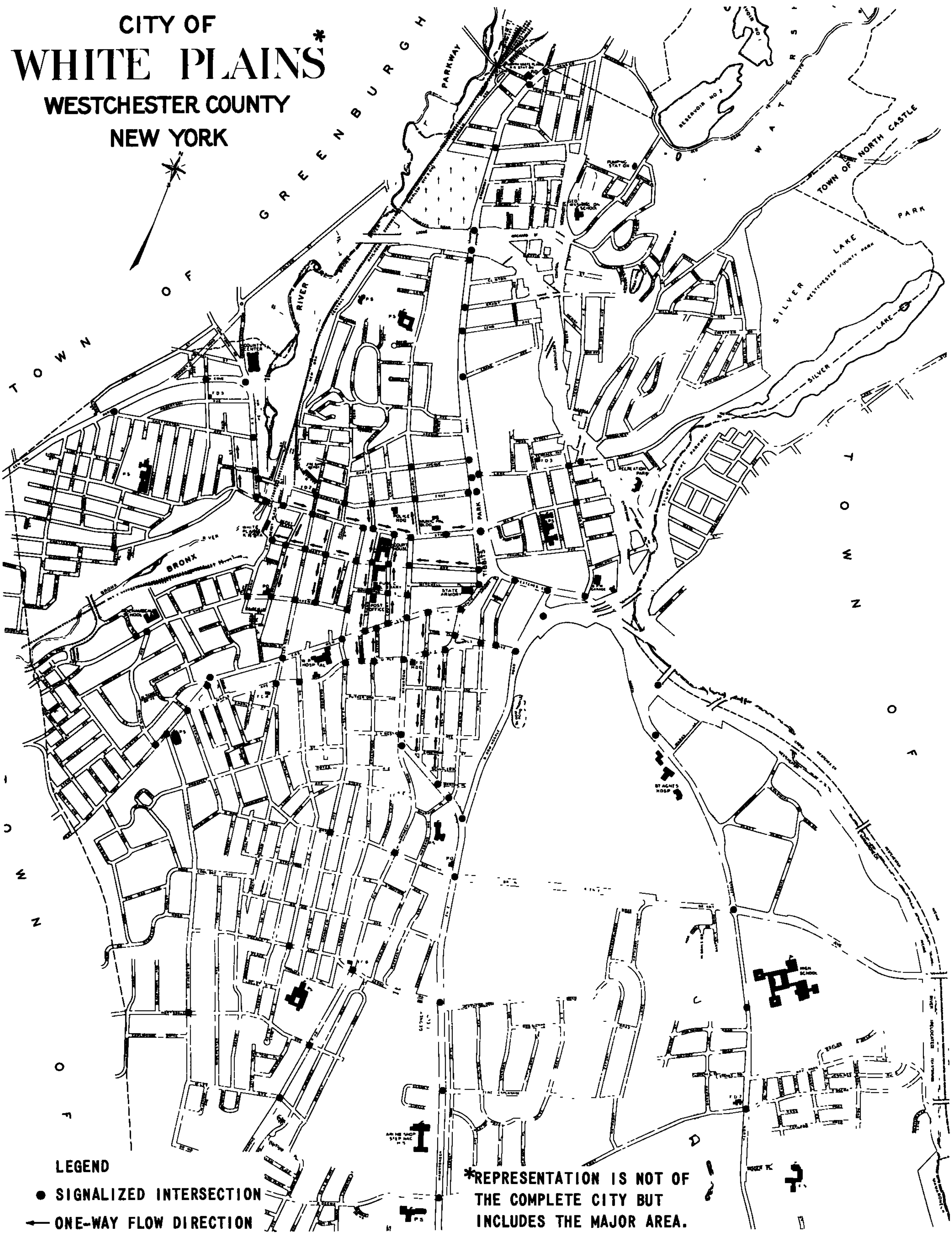


Figure 10. Locations of signalized intersections in White Plains, N.Y.

set of legs at a specific intersection. This lost time occurs during the portion of the amber interval in which no traffic flows through the intersection. The value of this lost time, $\tilde{\phi}'_{A|E,W}$, is predicted by forming the exponential weighted moving average of the measurements of $\phi'_{A|E,W}$ over past cycles. It should be noted that $\phi'_{A|E,W}$ is not directly observable with the proposed instrumentation, but rather must be calculated from a number of direct measurements.

Because the duration of the amber phase, $\phi_{A|E,W}$, is known, the calculation of the past $\phi'_{A|E,W}$ values should be performed as the difference between the total duration of the amber phase and that portion of $\phi_{A|E,W}$ during which the stop-line detector measures a non-zero vehicular flow.

The second example of arriving at the value of a predicted quantity concerns itself with departure flow rates, d_E, from the stop-line of an intersection.

One method simply uses the exponentially weighted moving average of departure rates, over corresponding intervals. These are referenced from the start of the green phase, $\phi_{G|E}$, and are taken over consecutive cycles. This prediction technique for obtaining d_E has the advantage of being both simple and yielding values which apply for a full cycle in the future. However, it is believed that better predictions of d_E may be necessary in certain applications to the equations.

Hence, a multiple linear regression, which linearly relates both the number of vehicles in queue and the measured departure rate during the previous interval, Δt, to the prediction of the departure rate during the subsequent intervals, is proposed.

It is believed that the general form of the prediction model would be necessary at intersections where left turns are permitted, with the associated conflicts.

APPLICATION TO A CITY

Inquiries were sent to a number of cities with populations of 50,000 to 100,000 within a 500-mile radius of Cornell Aeronautical Laboratory. A few replies indicated that the responding municipalities had acquired sufficient traffic data to be useful for this study. What was needed was a typical traffic model. Based on the physical characteristics, available data, and reasonable proximity to CAL, White Plains, N. Y., was selected as the sample city.

White Plains, New York, is 9.8 sq mi in area, has a population slightly in excess of 50,000 and a trade area for shopping of greater than 250,000, and currently has 116 signalized intersections (Fig. 10). By looking at each signalized intersection with respect to its distance from the closest signalized intersections, one-way streets, traffic movements, etc., subnetwork categories were determined as given in Table F-3.

It should be noted that each of the 116 signalized intersections was treated with equal importance; that is, the minimum-delay equations were applied, without prejudice, to each signal. As a result, it was estimated that 464 vehicle detectors would transmit information to the central site at any instant. The stop-line sensors were wired into the local controller so that they time-shared the detector electronics. The signal switching mechanism would activate only the sensors on the approaches facing a green signal.

SELECTING A COMPUTER SYSTEM

The data coming from each intersection are grouped into an array of 14 bits of information relating the status of each detector or sensor and the current signal state at the intersection. Also included is the intersection's identification number to assure that the information is used correctly. The input data format is shown in Figure 11.

When the computer program decides on the action to be taken with respect to a particular intersection, an output word consisting of 9 bits will be generated as shown in Figure 12.

Data Acquisition Requirements

The input word is presented to the computer at a rate of 60 samples per second, whereas the output word is generated, at most, once per Δt for each intersection. To determine if a new vehicle is approaching the intersection, a "01" combination is needed from the approach detectors, indicating that at the time of the previous reading there was no vehicle (0), and at the current reading a vehicle is sensed (1).

In the case of the departure detectors a combination of bits yielding "10" would indicate that a vehicle has departed from the stop line. When the approaching or departing vehicle is sensed by a detector, an appropriate storage location is updated by the computer. This technique yields a current vehicle count to be employed in the programmed calculations.

A decoding scheme, to be implemented by the computer, has been devised. It is proposed that each possible combination of the detector information be stored in the computer. From the eight bits containing the approach detector information, 256 possible combinations exist, of which 175 indicate at least one vehicle approaching. The

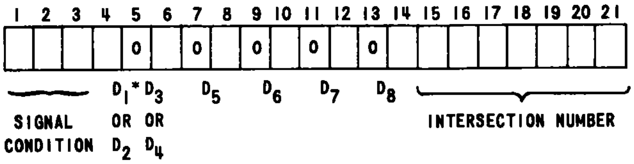


Figure 11. Input data format.

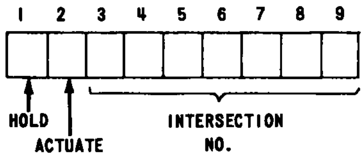


Figure 12. Output data format.

two stop-line detectors use only 4 bits of data, which give 16 possible combinations.

The decoding scheme will not require more than 76 computer cycles per intersection to decode. This number was arrived at by programming, for the available computing system, the case where a vehicle is detected simultaneously at each detector.

Based on calculations for 116 intersections and the 60-samples-per-second scanning rate, the time per computer cycle for data acquisition was found to be 1.89 microsec per cycle. Associated storage requirements were allocated as follows:

116	for the previous $D_5 - D_8$ detector data
116	for the previous $D_1 - D_3$ detector data
116	for the previous $D_2 - D_4$ detector data
928	for the vehicle counts at each detector
400	for the computer program (approx.)
335	for contingency and/or expansion
2,011	total for data acquisition

Delay Calculation Requirements

To evaluate the computer size and speed for the computations to be performed with respect to the minimum-delay equations, three basic expressions were defined as follows:

T_1 = estimated delay saved by those vehicles served (from the two legs) during Δt, which otherwise would have to suffer the amber, red, and lost time resulting from an immediate change in phase;

T_2 = delay saved which accrues to those vehicles served earlier in time as a result of the previous departures; and

T_3 = additional relative saving within the subnetwork, expressed as the sum of the individual savings at each of the neighboring signalized intersections.

The expressions T_1, T_2, T_3 were programmed in FORTRAN IV (*9, 10*).

By using 10 cycles for a multiplication and 18 cycles for a division (*8*), the total number of required cycles was computed for each equation, with the following results:

T_1 = Equation Type 1, 155 cycles
T_2 = Equation Type 2, 565 cycles
T_3 = Equation Type 3, 822 cycles

Since White Plains, N. Y. was used as the typical city, the 116 signalized intersections were categorized into 5 different types. Table F-3 shows the types of intersections with the proportion of each, and the size of the "saving" and "delay" equation necessary at each type, together with the required number of computer cycles. The results indicate that the total computational scheme will require approximately 900,000 cycles.

Also estimated were the storage requirements for the data that will have to be stored during the calculations (measured departures, predicted departures, measured arrivals, predicted arrivals, queue lengths, length of the signal phases for each intersection). These requirements totaled approximately 30,000 words of core plus 20 percent for contingencies and program instructions. Thus, approximately 37,000 words of core storage are needed.

The total number of calculations in the foregoing analysis must be performed within Δt seconds. If Δt is assumed to be 2 sec, the performance of 900,000 cycles within 2 sec implies a cycle time of 2.22 μsec.

Summary of Combined Requirements

The analysis results in the following requirements to achieve the projected data rate and computational schemes:

1. For the computational scheme, a computer should:
   (a) Perform 900,000 cycles in 2 sec (i.e., have a computer cycle time of not more than 2.2 μsec).
   (b) Have floating point capabilities (software or hardware).
   (c) Have address modification features (e.g., index registers or general registers).
   (d) Have a minimum of 37,000 words of core storage.
   (e) Have a data channel.
   (f) Have a real-time clock.
2. For the data accumulation scheme, a computer should:
   (a) Perform 76 cycles in 143.6781 μsec (i.e., have a computer cycle time of not more than 1.89 μsec).
   (b) Contain a minimum of 2,000 words of core storage (18-bit words).
   (c) Have address modification features (e.g., index registers or general registers).
   (d) Have a high-speed data channel.

Available Computer Systems

As a result of matching the specifications that were previously developed with off-the-shelf (existing) computer equipment, two possible systems were investigated in detail. The first is to acquire a single computer that can perform both sets of requirements sequentially in the specified time limit. The second system uses two computers—one for input-output and data storage, the other processing the data and making control decisions—operating in parallel. Details on cost analyses of the two schemes are given in Tables F-4 and F-5, respectively.

Both systems rent for about the same amount, $9,700 per month. Outright purchase of System A would cost approximately $412,000, and would require a monthly outlay of about $500 for maintenance. The costs for System B would be about $358,000 and $1,315, respectively. Over a 5½-year period, both systems would total about the same expenditure, $445,000, or $6,740 per month. The cost for the programming effort, to get the system into operation, is estimated at $25,000.

Additional Utilization of Computing System

Aside from doing the traffic controlling, either computer system could be used for the city's business and administrative data processing requirements. This can be done during a part of the day when the traffic is light and control can be handled satisfactorily by the local signal controllers.

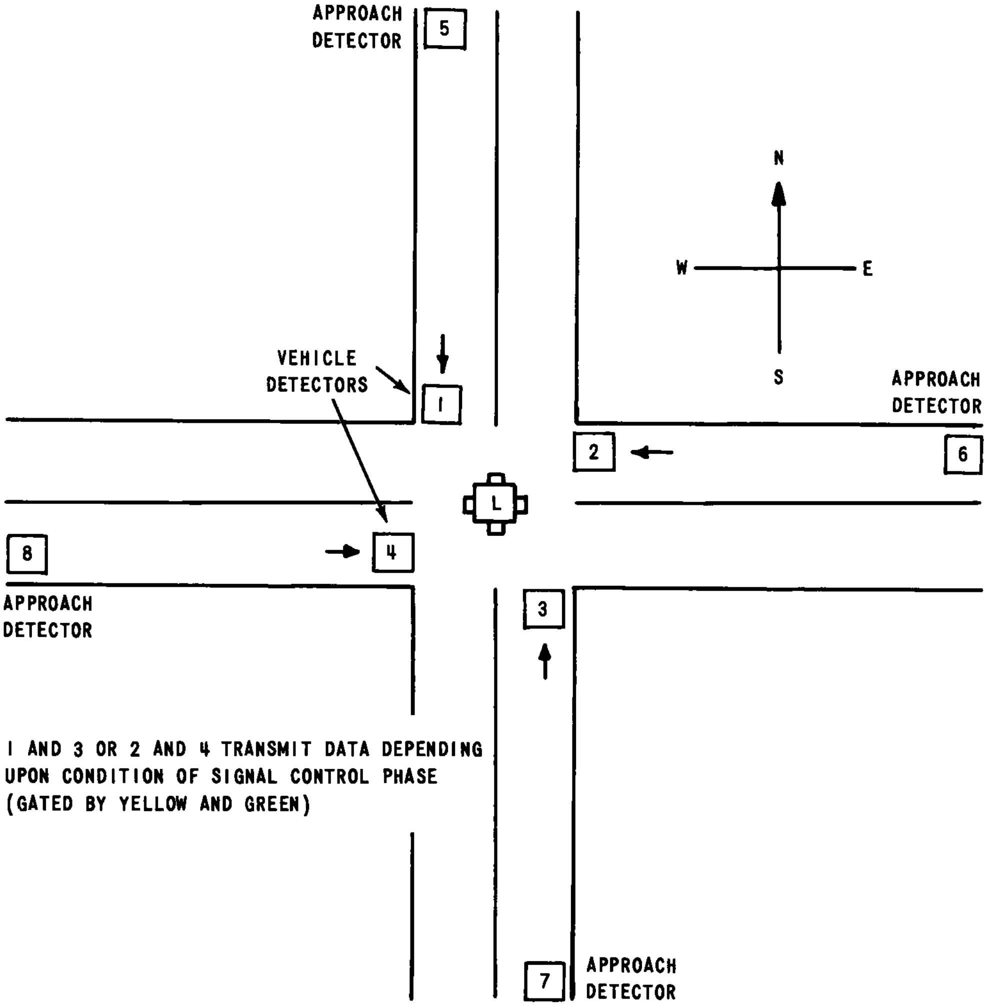


Figure 13. Intersection configuration.

In addition, the computer may be employed as a tool in the post-analysis of traffic data, as collected throughout the control periods or at other specified times, in order to improve its control function. It may also be used in the performance of accident studies, etc.

INSTRUMENTATION

Figure 13 shows an intersection with eight vehicle detectors —four located at the stop lines and four several hundred feet upstream on the approach lanes. This represents the maximum amount of detection equipment required at a single intersection. The sensors proposed are of the induction loop type, which initiates detector contact closure when a vehicle passes over the loop. To minimize the amount of detector electronics necessary, a time-sharing technique would be employed whereby the two detectors for the stop line would be switched between the four stop-line loops, depending on the phase of the signal. The approach detectors would remain active at all times, thereby providing continuous data on approaching traffic. The logic circuitry necessary to accomplish the switching function has been designed. The instrumentation for the detection of the state of the signal, together with a phase pattern and coding table, is shown in Figure 14. Only conditions of three bulb circuits (red and green circuits along one line of travel, and green circuit for the crossing line) are needed for determination of signal state.

Input Logic

The data from each intersection must be transmitted to a control site where the computer is housed. Because all information is of an on-off (binary) nature, a selective signaling system using telephone lines could serve as a transmission medium. Each channel of information would have an assigned signal frequency. At the central site tuned receivers would be used to detect the information. Each telephone line would then carry a composite of tone signals in both directions, between intersections and the central (computer) site. The data from each intersection would be decoded and grouped to form a computer word whose structure is defined in Figure 11. Data from each instrumented intersection are brought to the central site con-

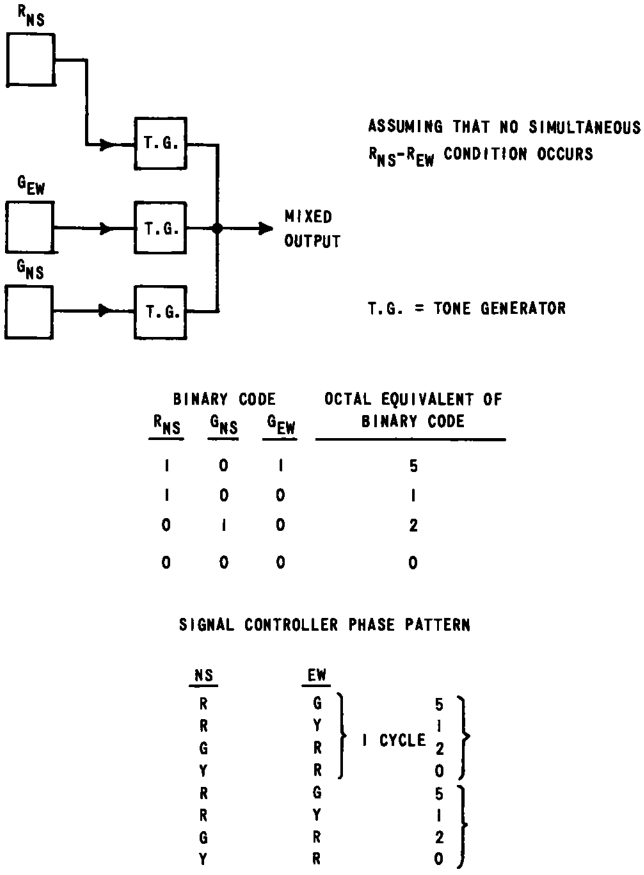


Figure 14. Monitor for phase condition of signal controller.

tinuously, whereas the computer control commands are sent back to each intersection only as required. Figure 15 is a logic diagram of the input circuitry.

Output Logic

The output control signals from the computer are generated at a maximum rate of one every Δt seconds for each intersection. They consist of a HOLD and an ACTUATE command. When it is desired to transfer a signal at a particular intersection from fixed-time operation (local controller) to computer control the output word presented by the computer will contain a binary "1" in the HOLD bit and the corresponding 7-bit intersection identification. Once this has been accomplished, if it is desired to change the signal state the ACTUATE bit must be transmitted as a binary "1". Figure 16 shows a logic diagram of the output circuitry.

SYSTEM INSTRUMENTATION COST

An investigation was made with reference to the sources of supply for the instrumentation required at each intersection and at the central site. Based on the mode of operation desired, a typical set of equipment cost figures was compiled for the complete installation. The breakdown was as follows:

Intersection equipment, @ $1,575	$182,700
Input logic interface at central site	26,530
Output logic interface at central site	19,740
Telephone line leasing	2,000/mo.
Total	$228,970 plus $ 2,000/mo.

These figures are approximate, as telephone leasing rates will depend on the community selected. Not included are the maintenance and construction costs associated with the equipment's installation and upkeep.

CONCLUDING REMARKS

There is little doubt that a traffic control system such as the one described in this report can lead to improved traffic movement in an urban complex, when compared to a system that is not traffic responsive. Investigations (*3, 4*) have shown that this improvement is purchased at a lower cost than the penalty paid for doing without it.

There are, however, signal control systems that are responsive to the demands of traffic, to a greater or lesser degree, that do not use a general purpose computer for deciding on signal settings. These systems operate on preprogrammed control modes that are called into effect by measurements at a point, or several points. The great majority of the controllers use a special purpose analog computer to discriminate among levels of traffic demands in order to call for changes of modes. Flexibility of adapting the control mode to suit the traffic situation is limited by the finite number of modes that are physically wired into the controller. In most cases the cost of these control systems is more than offset by the value of the reduction in motorists' aggregate travel time.

The latest entry into the field of traffic control is the high-speed general purpose digital computing machine. A control system using this machine also is preprogrammed. Unlike the special purpose computer, however, its flexibility to meet changing traffic requirements is not established at the factory, although the storage capacity and computing speed, which are determined by the combination of hardware and software components are, of course, limiting physical factors of the machine. Conceptually, the control doctrines that may be programmed into the computer are limited only by the ingenuity of the analyst who develops the controlling equations and modes. As current and future hypotheses are tested and become candidates for practical control applications, the computer generally will be unlimited in its adaptability to implement these control doctrines. Moreover, these changes are effected without adjustments or alterations to the hardware in the field, but simply within the software (the written program and the data input) at the computer site. As one traffic engineer stated the case: "The general purpose computer gives me an easy way to correct the mistakes I am going to make, because I *don't* think I know everything that the traffic is going to do in my city."

The justification for considering any traffic control sys-

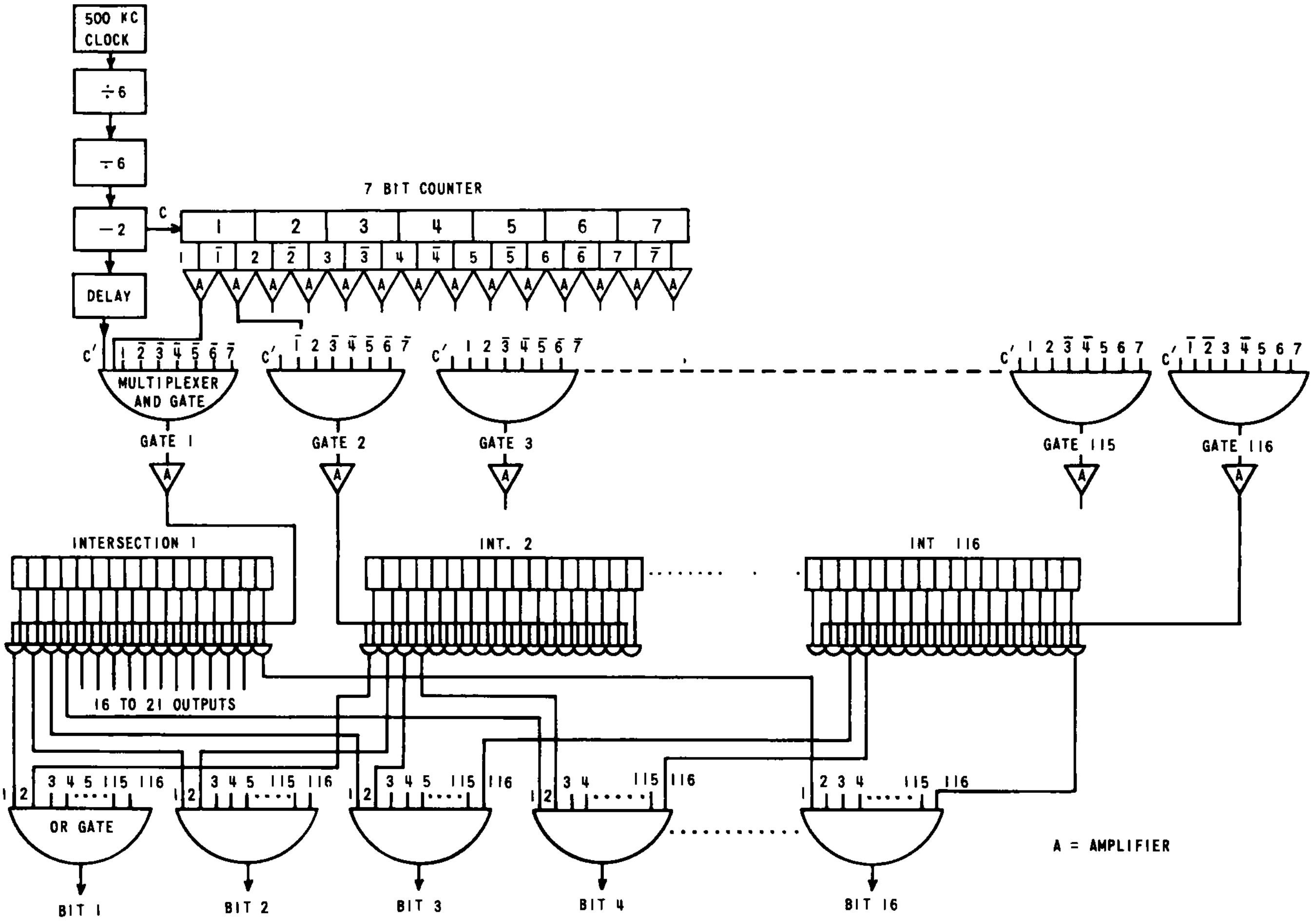


Figure 15. Computer central site input control logic diagram.

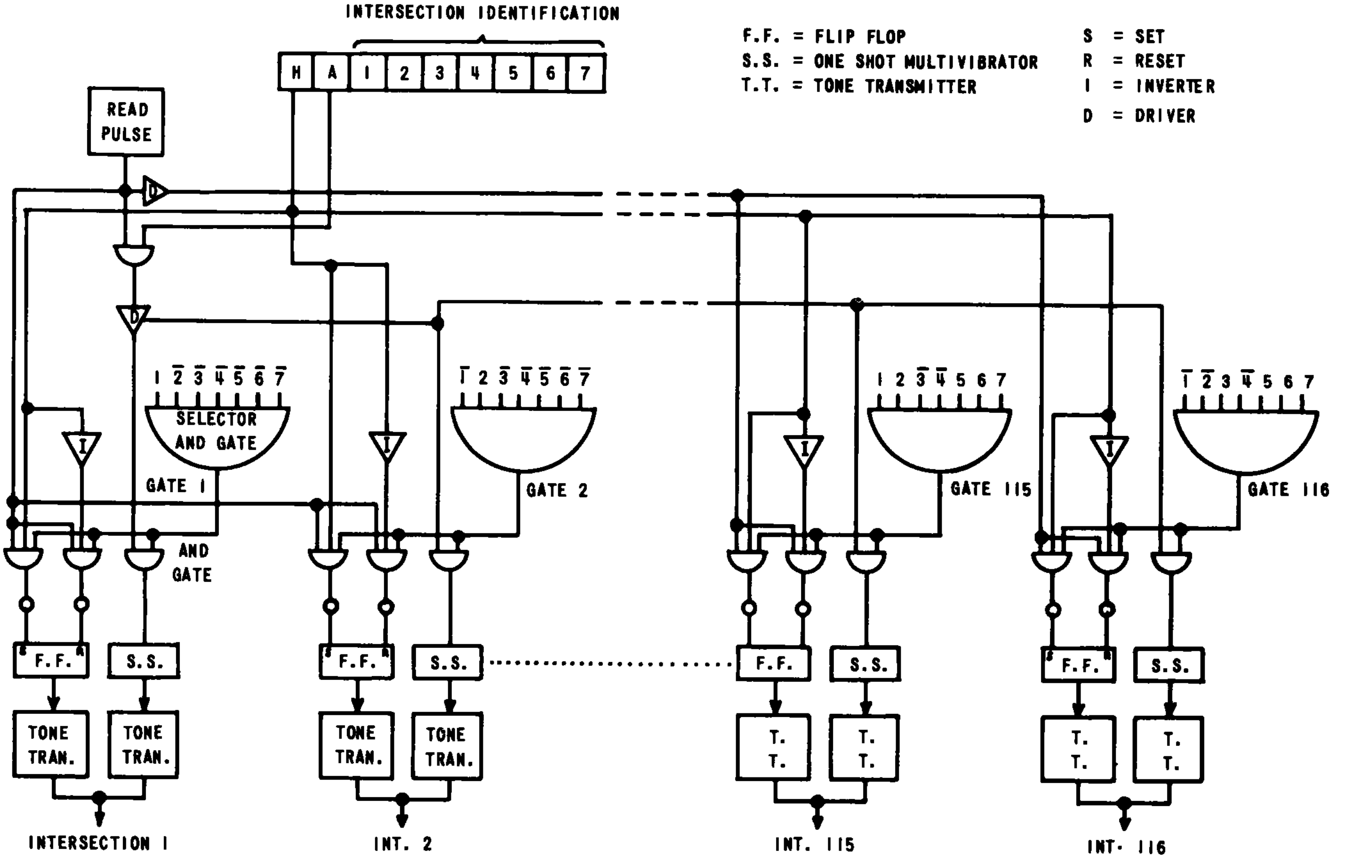


Figure 16. Computer central site output control logic diagram.

tem must rely on at least two basic premises. First, there is a current or future need for improvement; and second, the system appears to offer a practical solution when considering the alternative courses of action. Although improvements in traffic flow can be achieved by increasingly efficient assignment of right-of-way, the level of service must approach, as an asymptote, conditions at saturated flow. When the demands continue beyond this point, either the service must deteriorate or the inexorable alternative of "laying more concrete" must be followed.

The original problem reduces to that of determining the increment of service that may be obtained now and in the future by such an installation. Proverbially, this may be stated: "How much can my investment buy for me?" The question of what a digital-computer-controlled system may accomplish now is being evaluated by researchers at Toronto, Canada; San Jose, Calif.; and Glasgow, Scotland. The future is more difficult to analyze. It is for this very reason that the digital computer deserves still further consideration.

Finally, it is in order to comment on the "minimum-delay" control mode that has been developed in this research. Of the three digital computer systems currently being evaluated, only the Glasgow program has definite plans to include a control mode that uses minimum delay as a criterion. The others are testing modes that derive principally from the more conventional schemes of the past.

REFERENCES

1. WEINBERG, M. I., "Traffic Surveillance and Means of Communicating with Drivers — Interim Report." *NCHRP Report 9* (1964).
2. WEINBERG, M. I., ET AL., "Surveillance Methods and Ways and Means of Communicating with Drivers." *NCHRP Report 28* (1966).
3. CASCIATO, L., and CASS, S., "Pilot Study of the Automatic Control of Traffic Signals by a General Purpose Electronic Computer." *HRB Bull. 338*, pp. 28-39 (1962).
4. MILLER, A. J., "A Computer Control System for Traffic Networks." *Proc. Second Internat. Symposium on Theory of Traffic Flow*, London (1963).
5. GREENSHIELDS, B. D., ET AL., "Traffic Performance at Urban Street Intersections." *Tech. Rep. No. 1*, Yale Bureau of Highway Traffic (1947).
6. BLUMENTHAL, R. C., and HORN, J. W., "Pressurized Intersections." *HRB Bull. 167*, pp. 42-62 (1957).
7. KELL, J. H. "Results of Computer Simulation Studies as Related to Traffic Signal Operation." *Proc. 33rd Ann. Meeting*, Inst. of Traffic Engineers (1963).
8. "IBM 7040/7044 Principles of Operation." Form A22-6649-4, IBM Systems Ref. Lib. (May 1964).
9. "FORTRAN II General Information Manual." Form F28-8074-3, IBM Systems Ref. Lib. (Dec. 1963).
10. "IBM 7040/7044 Operating System (16/32K) FORTRAN IV Language." Form C28-6329-2, IBM Systems Ref. Lib. (1964).
11. HOLT, C., MODIGLIANI, F., MUTH, J., and SIMON, H., *Planning Production, Inventories, and Work Force.* Prentice-Hall (1965).
12. GREENSHIELDS, B. D., and WEIDA, F. M., *Statistics with Applications to Highway Traffic Analyses.*" The Eno Foundation for Highway Traffic Control (1952).
13. WOLD, H. O. A., "Unbiased Predictors." *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 719-761 (1960).
14. ROSENBLATT, M. (Editor), *Time Series Analysis.* Wiley (1963).
15. MOOD, A., *Introduction to the Theory of Statistics.* McGraw-Hill (1950).
16. ARLEY, N., and BUCH, K., *Introduction to the Theory of Probability and Statistics.* Wiley (1950).
17. KELL, J. H., "Intersection Delay Obtained by Simulating Traffic on a Computer." Inst. of Transportation and Traffic Eng., Univ. of California, Berkeley.
18. City of Buffalo, N. Y., "Specifications for Master Computing Equipment for Two Interconnected Traffic Control Systems on Main Street (North System and South System)."
19. Vehicle Traffic Control Systems, General Railway Signal Co., Rochester, N. Y., "Traffic Control and Surveillance Computer—Main Street Traffic Control System, Buffalo, New York."
20. MATSON, T. M., SMITH, W. S., and HURD, F. W., *Traffic Engineering.* McGraw-Hill (1955).
21. "Research on Traffic." British Road Research Laboratory, HMSO (1965).
22. HORTON, T. R. (Editor), *Traffic Control, Theory and Instrumentation.* Plenum Press (1965).
23. "Two-Phase Full-Vehicle-Actuated Electronic Traffic Controllers." FVA-210 Series, Crouse-Hinds Co.
24. "Semi-Vehicle-Actuated Electronic Traffic Controllers." SVA-110 Series, Crouse-Hinds Co.

APPENDIX A

VEHICLE DETECTORS AND THEIR LOCATIONS FOR SURVEILLANCE OF TRAFFIC IN A STREET NETWORK

The traffic signal system under study is to be activated by commands from a high-speed digital computer operating in real time. Surveillance over the movement of traffic, in response to the control system, is to be supplied by vehicle detectors located so as to provide the data required by the control doctrine. The following is a discussion of the types of detectors that are appropriate for this purpose, the ways in which the sensor outputs may be processed, and the number and locations of sensors compared to the information they might furnish.

SENSOR TYPES

Synthesis of this digital-computer-controlled system has been restricted to the use of "off-the-shelf" (currently available) equipment so that the descriptions of the detectors are of commercially available items. Later there is a treatment of the manner in which the output from one of the sensors might be processed in order to increase the amount of information, over that presently provided, that can be furnished to the computer. The detectors are of the following types:

1. Overhead pulse-type presence detector.
2. Overhead continuous wave (CW) Doppler velocity detector.
3. In-pavement magnetic loop presence detector.

The overhead pulse-type presence detector is used extensively in traffic signal control and, to a lesser degree, for the acquisition of data on traffic flow. It uses time-lapse discrimination to distinguish between return of a pulsed super-audio acoustic signal from either the pavement below the detector or a vehicle over the echo "spot." Because of time required for the pulse to traverse the usual 36 ft from transmitting to receiving transducers (18 ft above the pavement) in cold weather, the pulse repetition frequency (prf) is limited to about 25 cycles per second. This discrete limitation determines the accuracy with which the time can be resolved that a moving vehicle spends under the detector, a point that is discussed in some detail in NCHRP Report No. 28 (*2*). Vehicle speed is estimated by dividing a previously statistically derived average length of vehicle by the time spent under the detector. The latter quantity is computed as the number of "vehicle presence" returns divided by the prf. These detectors are fairly reliable and cost something over $500 each plus the cost of a supporting structure for the sensory device.

The CW Doppler velocity detectors use both acoustic and radio frequencies for signal radiation. Both types measure Doppler frequency shift to indicate the velocity of an approaching vehicle. The bandwidth of the frequency filters determines the resolution of the speed indications. Practical applications have shown that the accuracy of the detectors falls off at low speeds, beginning at about 20 mph and continuing down to the lowest speed at which they seem to operate, about 4 mph. The cost of these detectors is similar to that of the overhead presence detector.

The in-pavement magnetic loop detector senses the presence of a vehicle by the shift of phase in an alternating-current circuit in which the inductance of the loop and a capacitance are tuned to oscillate at a reference frequency. The phase shift causes a voltage pulse that is used to activate a switch, which can remain closed until the vehicle leaves the loop and the circuit retunes. Because the phase shift is affected by the mass of the vehicle and its proximity to the loop field, there can be significant variations between the leading edge of the vehicle when it first enters the loop and its location when the loop has recorded a detection. Similarly, this variation can exist when the vehicle leaves the loop.

As a result, the time required for a vehicle to pass over the detector loop may be measured with a significant error, some of which may be biased out by correlation of detector output and observation of speed and vehicle length by very precise methods, as can be done in the case of overhead presence detectors. Prices for loop detectors vary from $120 to more than $500.

Each of the detectors described will furnish a digital count of each vehicle that activates it.

SENSOR DATA PROCESSING

The sensors used in the vehicle detector systems provide raw data that could be processed in ways different than they are. For instance, the overhead presence detector uses a logic circuit that closes a switch when, after a period of time that spans two pulses, the echo time for the pulse is shorter than some selected value. Thus, when two successive pulses indicate "vehicle" presence it is assumed that a vehicle is present with negligibly low probability of "false alarm" (spurious signal). Similarly, the switch is opened when this same logic is satisfied for the "no vehicle" indication. Another version of this detector, using the same sensor and costing about 50 percent more, is able to discriminate the time delay for the echo at a second level, thus providing ternary information. It classifies vehicles as "high" or "low," as well as determining that there is no vehicle.

It is evident that the time delays could be measured on an analog basis, if so desired. The result would be a set of echos that would establish the "profile" of the vehicle that passed under the detector, the resolution of the profile being dependent on the vehicle speed, for a given prf. Such a profile might be used to identify the type of vehicle much more closely than is afforded by the "high" or "low" classification. The technique that might be used to match the acquired profile with some known profile is not discussed here other than to comment that it could be done by any of several signature recognition methods (map matching, Fourier analysis, etc.)

The CW Doppler sensors used as velocity pick-ups transmit and receive "tones." Unless their designs were to be changed quite radically, they are limited pretty much to their present performance. For instance, if the antenna were to operate as a "seeker," as in the case of homing radars, it would be possible to determine vehicle speeds quite accurately from the rate of rotation of the antenna at some specified angle.

The phase shift or voltage change sensed by an in-pavement loop detector is a continuous function that can be recorded as a profile or "signature." These raw data can be more informative than those acquired from the discrete prf of the overhead detector, because, in theory, the completeness of the signature is not affected by vehicle speed. If a statistical investigation of these signatures indicated that a high level of confidence could be placed in using them to classify vehicles, they could be used to increase greatly the amount of information going into the computer storage. Not only would the types of vehicles be known, but better estimates of individual vehicle speeds also could be computed because their lengths could be determined more accurately. Again, any of the usual techniques for pattern recognition could be employed. A really powerful discrimination logic might make it possible to separate and identify two signatures simultaneously sensed by a loop covering two traffic lanes.

For the time being, the advantages to be gained by additional processing of the sensor outputs are considered only for their academic interest. The control system that will be designed will use only the current output of the instrumentation, which is merely binary information on a switch position, not the signature profile suggested in the foregoing.

DETECTOR LOCATIONS

Two categories of detector installations on each approach leg to an intersection are discussed. In one of these, one detector location is assumed for each approach leg; in the other, two detector locations are assumed. Because of the use to which the information on the traffic movements is to be put and the limitations in accuracy noted for the velocity detectors, the employment of presence detectors is assumed. At this time, no comment is offered as to a choice between the overhead and loop-type sensors.

One Detector on Each Approach Leg

For simplicity, an intersection of two 2-lane streets is used, with each street carrying two-way traffic. In this configuration it is possible to have the detector located at the stop line or some arbitrary distance upstream. What might happen in each of these cases is considered.

If the detector is at the stop line, the only information that can be elicited during the red phase is that there is none or one vehicle waiting for passage. Also, there can be no assurance that counts from upstream can be used to foretell queue length unless turns are prohibited. The condition is too restrictive, making a detector location at the stop line relatively poor. As shown in Figure A-1, movements at intersection 2 will not reveal the number of vehicles moving toward intersection 1, except on a statistical basis, because turns and through movements at intersection 2 cannot be resolved.

Therefore, if it is decided that the detector must be upstream of the stop line, the question is, "How far?" A fairly obvious question to be resolved is the length (number of vehicles) of a queue waiting for a green signal. If it is desired to estimate the number of vehicles in the queue that will begin to form when the signal turns red, "how far" can be expressed in stopped vehicle lengths (about 23.5 ft for passenger cars). Intuitively, it would seem that the higher the volume handled by the intersection, the longer should the signals be phased, implying longer queues (for isolated intersections) and, therefore, a large distance from the intersection to the detector.

Lane volumes for city streets in which intersections are controlled and time-shared appear to reach 1,100 vph of green time (*6*, Fig. 11). In his mathematical study of the effect of variation of cycle lengths and splits in a fixed-time controller, Kell (*7*) commented that cycle lengths of 30 sec were about as short as could be tolerated for lane volumes of 500 vph. Optimum cycle length for equal volumes of 500 vph on each approach, with no turns permitted, was 50 sec. "Optimum" means that average delay for each vehicle was minimum, at 17 sec. Disregarding the short amber phase, either 7 or 8 vehicles would pass through the intersection during the 25-sec green phase.

If 8 vehicles are waiting in queue for the signal, the detector would have to be not less than $8 \times 23.5 = 188$ ft from the stop line to have counted these vehicles. (To make this count the surveillance system and computer logic must be able to identify the lead vehicle of the stopped platoon, a point discussed later.) If the volume is less than the maximum, and the lead vehicle can be identified, the number of vehicles in the queue can be counted with reasonable accuracy. However, if for some reason the queue exceeds the length to the detector, there can be no knowledge of the location of the upstream end without additional information from some upstream location.

An absolute upper limit to the distance of the detector from the stop line can be estimated from the results of Kell's work. At 700 vph on each approach lane he reported (*7*) that cycle lengths of less than 150 sec failed, "failure" being defined as a queue length that exceeded 100 vehicles and average waiting time in excess of 10 cycle lengths. All cycle lengths failed for volumes above 700 vph.

Assuming the foregoing volume and cycle length, 29 vehicles would have to pass through on each green phase, giving the longest meaningful queues that might be imagined for an operational system. The length required to store 29 vehicles is 685 ft. This distance is unrealistic; it is unlikely that a signal would be permitted to remain on a 2.5-min cycle at an ordinary intersection. Also, a volume

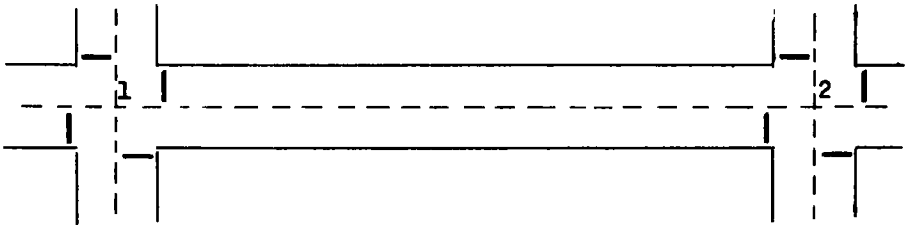


Figure A-1. Detectors at stop line.

of 700 vph per lane is academic; real values would be more like 500 to 550 vph. Therefore, it seems that a reasonable location for a detector at a busy intersection would be about 200 ft back from the stop line, unless it was observed that queue formations regularly exceeded this distance. Another consideration might be the control logic's need for some minimum time interval between a vehicle's passing the detector and its expected arrival at the intersection.

Summarily, a desirable location for the detector can be determined from observed traffic flows at an intersection. The set-back should be great enough to include the longest queues that are normally expected to form. As this distance is increased, however, the probability of identifying the vehicle that will be the lead car in the queue will decrease.

Identification of the lead vehicle can be done from a statistical study by observing the average time lag from passage over the detector to start of the amber phase for the vehicle that becomes the platoon leader. It should be noted that there is no way to identify the vehicle positively, short of complete surveillance on the approaches. At this time it is not economically feasible to provide the instrumentation required. Figure A-2 pictures the situation just as the amber phase has begun; the time is $t = t_a$. From long-term observation it has been determined that when $t_a - t_d \geqq K$, the vehicle passing the detector at t_d will go through the signal. The situation in Figure A-2 is: $t_a - t_{d_2} > K$, so vehicle 2 goes through; $t_a - t_{d_3} < K$, so vehicle 3 will be at the stop line and there will be at least three vehicles in the queue. If vehicle speeds are sensed, the value of K can be made speed-dependent.

Alternatively, a subroutine can be written into the program logic that can use frequency of arrivals over the detector as a clue to average speed of the traffic stream. Such a relationship could be unambiguous if it is found that the movement of traffic on city streets is always below the critical speed, v_{crit} (*2*, App. A), on the volume-density curve that describes the traffic flow. Also, it is possible that the value of K might be affected by whether or not the identified lead vehicle really is at the head of a moving platoon, in which case its movement would not be impeded by previous traffic. This can be determined by examining the time gap that preceded the vehicle signal that is later identified as that of the leader.

Two Detectors on Each Approach Leg

Because a suggestion to use more than one detector on each approach to an intersection immediately prompts the question, "Why," the discussion must center on what additional detectors, and how the benefits to be derived compare with the additional cost.

It was stated previously that a presence detector at a stop line would indicate with certainty that there was a vehicle waiting for passage, and also could count vehicles passing. But the stop-line detector has no way of supplying information on the length of a queue, which is an important factor in determining a control decision. It does, however, lend itself to instrumenting a computer logic to decide if a vehicle is waiting for an opportunity to turn left. For, if a vehicle is "holding" at the intersection, as shown by the detector, and a steady stream of counts is coming in from the detector on the opposite approach, it may be surmised that the holding vehicle is waiting for an acceptable gap

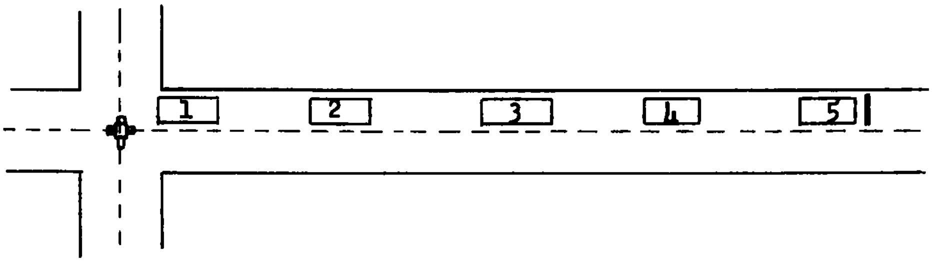


Figure A-2. Mid-block detector.

in order to turn left. Any other "hold" signal would not be normal during green time.

A persisting presence signal on a stop-line detector could indicate some sort of trouble, if it continued for a protracted period in the face of no sign of conflicting traffic. Although a similar logic can be written for a mid-block detector location, which should be just as effective during periods of high traffic activity, it will involve a certain amount of additional delay time before deciding that some difficulty exists.

A stop-line detector also will lend strength to the logic described previously for identifying the lead vehicle of a platoon stopped at a signal. It is evident that an indication of a stoppage at the intersection will invalidate the scheme described in connection with Figure A-2. On the other hand, if the passage of vehicles at the stop-line and mid-block detectors reaches a fairly steady state during the green phase, this is reasonable assurance that the logic will hold. Actually, it is necessary only that an indication of vehicle departures at a satisfactory frequency be obtained or that no further vehicles are attempting to move through the intersection.

If the stop-line detector indicates a blockage and the identification logic is voided, it is still possible to compute the number of vehicles in the waiting queue if, during the red phase, the queue extends to the mid-block detector. The loop must be long enough so that it will not lie within the gap normally expected between vehicles in a queue; if there is occupancy to or past a detector, a presence must be shown. This latter criterion must hold for the stop-line detector as well. The 23.5-ft spacing of vehicles in a queue is based on an average vehicle length of 17.5 ft and an average gap of 6 ft. Thus, it may be suggested that a loop length of 10 to 12 ft, in the direction of travel, would be fairly certain to detect any stopped vehicles without losing discrimination between successive vehicles when they are in motion.

Once the number of vehicles in a waiting queue has been determined, the counts at the stop line can be subtracted and those at the mid-block detector can be added to keep a running inventory on the number of applicants for passage through the intersection. Although errors are certain to creep in, this method may be used over the course of a green phase with high confidence. Then, if the queue length can be computed by the means discussed previously, the data can be reinitialized. If not, the data can be extended through each signal cycle, with decreasing confidence, until reinitialization can take place.

Is there any advantage to a detector location somewhere between the stop line and mid-block? It is obvious that it will give a better idea of the lead vehicle of a queue than will the mid-block detector. A problem still will exist, however, in determining the identification of the times at which a vehicle has crossed each detector for the counts from the two detectors to be correlated to give numbers in the queue. This identification should occur without error, if the use of the second detector is to be justified. Identification without error may be hard to achieve in city traffic, considering the uncertainties due to acceleration noise. If the mid-block detector uses a velocity sensor, it may be possible to get an expected time of arrival by extrapolating detected speed over the distance separating the detectors. However, this would be done at the expense of an indication of presence at the mid-block detector. The assumption is made that a presence detector cannot be used to compute speed of a particular vehicle with satisfactory accuracy.

Most important of all is that dual detectors upstream of the stop line are justified only if they enable the control system to make consistently better estimates of traffic conditions than will a single detector, and that the savings in delay time that may result shall more than pay for the over-all cost of the additional detectors. (This additional cost is not in the detectors alone, but must include increased capacity within the computer for data storage and computational logic.)

There appears to be no particular virtue to a stop-line-and-far-side installation, as shown in Figure A-3. If the streets are two-lane, it should be possible to get information on turning movements, using an appropriate computer logic. For standard city block spacings, the far-side location probably would be too far upstream for the ideal mid-block location, particularly where parking is permitted.

A comment is offered about the use of stop-line detectors in cities in which buses operate. It is possible to include the effect of bus stops in the control system if a beacon in the bus will radiate a special tone. A scheme of this sort could be used to denote when a bus is approaching an intersection or is discharging or taking on passengers. The loop, acting as an antenna, would pick up the transmitted signal from the bus and carry this information to a filter circuit, where the bus "signature" would be recognized and the computer so informed.

Addition of stop-line detectors at a normal intersection (Fig. A-1) can be done by using four sensors coupled in pairs to two detector electronics packages. It is evident that no traffic should be passing the stop line on the red phase, so information from the sensor on that approach should be trivial. Only when the green light is shown is the movement, or lack of it, of interest to the control system. Therefore, an economy in hardware can be effected by providing for automatically switching the sensor to the detector when the green phase is in operation. Because information on the state of the signal will be sent to the

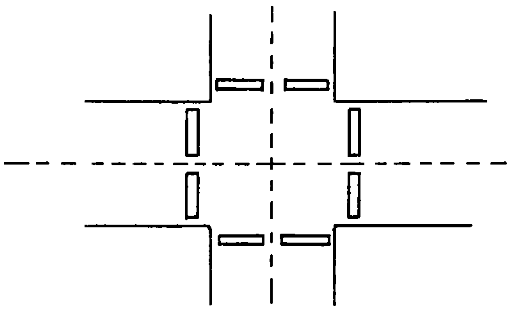


Figure A-3. Stop-line and departure detectors.

computer, the data from the sensor can be sent to the address in the computer for the proper approach leg.

Although no particular example is cited, it can be seen that a single detector electronics unit can be time-shared by several stop-line sensors for control of a polyphase signal cycle. This is possible so long as the state of the signal is being sent to the computer, so that the data can be processed with proper meaning.

The traffic data to be derived by the surveillance equipment will be used to provide for the solution of a minimum-delay equation, which is discussed in a preliminary manner in Appendix D. Information supplied by stop-line detectors can be used directly for vehicle departures, without having to accept an estimate and the time lag that would result if the data come from mid-block detectors.

CONCLUDING REMARKS

The preceding discussion has been concerned with location of detectors at an isolated intersection of 2-lane streets. Many digressions from this type of intersection occur in a real street network. Thus, it is to be expected that the numbers of detectors required to serve the system adequately may vary considerably from intersection to intersection.

APPENDIX B

AN ESTIMATE OF TRAFFIC SURVEILLANCE AND SIGNAL STATE DATA FOR A DIGITAL–COMPUTER–CONTROLLED TRAFFIC SIGNAL SYSTEM

This appendix describes and discusses some of the information that will be transmitted from the traffic signals and the surveillance equipment. The incoming data will have to be coded, transmitted, stored or read directly into the computer, and then processed by the computer in order to arrive at control decisions.

ESTIMATE OF TRAFFIC SURVEILLANCE INPUT DATA

In addition to developing the "software" for the control system, it is planned to synthesize the hardware elements, over and above the actual traffic lights and local signal controllers at each intersection. To do this, a number of decisions must be made concerning the equipment. For instance, the average number and types of traffic sensing detectors must be estimated, assuming that detectors are installed at each signalized intersection. The number of data communication and signal control lines needed will depend on the type of sensors used and their information rates, as well as the degree of flexibility designed into the control logic.

The City of White Plains, N. Y., was selected as having a well-documented traffic "model," which was used to determine the requirements for the system. The city currently has 116 signalized intersections, some operating independently and others working in groups to provide coordinated control.

It is assumed that there will be an average of 4 detectors at each signalized intersection, or a total of 464. If each detector is used to record vehicle counts on one lane, the frequency of interrogating such a detector will be different than if it is used for surveillance over more than one lane. (In Appendix A it is suggested that the most suitable type of detector for this application is the magnetic loop installed in the pavement.)

The loop detector is able to distinguish between nothing in its field and something that produces enough of a change in the field so that the detector circuitry can sense the voltage change or phase shift. If the "something" produces a sufficiently large signal-to-noise ratio of a reasonably consistent value, it is possible to design the detector so that it can discriminate between one or two "somethings," possibly more. In general, the alteration of the field by a single vehicle is not a step function of constant value, so that positive discrimination reduces in probability as the number of lanes covered by the detector increases. In this study it is assumed that no loop will be installed across more than 2 lanes.

On a statistical basis, the loop detectors will be credited with the ability to sense the "length" of the vehicle with sufficient accuracy to provide a measure of lane occupancy. However, it is possible to set the signal-to-noise threshold level high enough so that leading or trailing portions of the vehicle will lie within the field and yet a null signal will be registered. This latter point applies particularly in the case of the multilane application. Here it is possible to mistake one vehicle for two, or vice versa, depending on the discrimination threshold levels and the amount and distribution of electromagnetic material in the vehicles.

To illustrate this latter point, Figure B-1 shows the hypothetical signature of a large tractor-trailer truck and a small imported vehicle. The traces show the voltage levels as the vehicles pass into and out of the zone of sensitivity of the loop. If the two vehicles reached the loop simultaneously, as shown, there might be a significant time lag before the loop output rose to the two-vehicle level. Also indicated is the possibility that the signal might fall to a low enough level along the length of the truck so that a small car also might be within the loop without its presence being known.

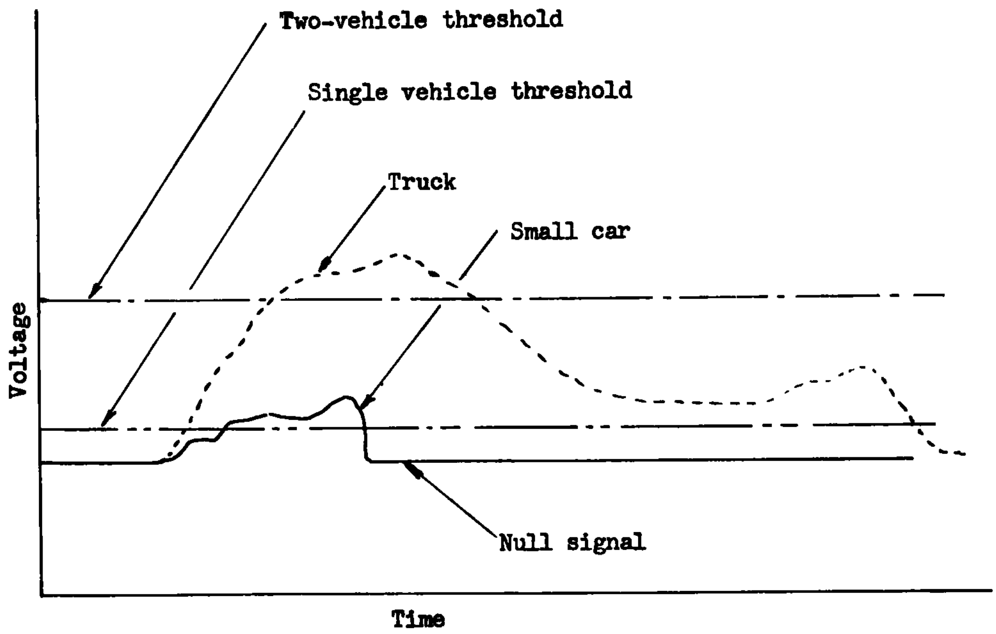


Figure B-1. Analog trace from loop detector.

Yet if the two-vehicle threshold is set low enough, the truck could look like two vehicles to the detector. Actual performance of the detector system must be based on the vehicle signatures. A decision on the accuracy of the ternary-level (2-lane) system must be held in abeyance until the signature traces are available.

Once the decision is made that the detector circuitry is adequate for 3-level discrimination (probably on a statistical basis) it must be determined what accuracy of counting can be obtained. It is evident that arrivals of vehicles over the loop can range from simultaneous to tandem. The solution to the counting problem would be simple and straightforward if the signatures were fairly sharp step functions and of reasonably uniform value, and if information from only one detector was being used.

But the signatures may not be step functions and many detectors will be feeding information to the computer. This results in a problem of time-sharing for the detectors, for which a frequency must be determined that will provide adequate resolution for sensing the arrival (and departure) of vehicles.

The lowest frequency at which any single detector for a single lane may be interrogated will be that for the shortest time period that a vehicle will be over the loop. A small car, perhaps 100 in. long, traveling at 60 ft/sec (40 mph), will pass a point in 0.139 sec (7.2 cps). At any but very low speeds, the length of the vehicle will be less than the space between vehicles, the latter being about 6 ft when a line of traffic is stopped.

When the loop is required to cover two lanes, the interrogation frequency requirement changes drastically. Suppose, as shown in Figure B-3, the signature after processing is a voltage level for two vehicles at close proximity but in different lanes. For the sake of simplicity, the signatures are shown as "square waves" in Figure B-2. It can be seen that, to determine that the signal is not at constant level from the front of the first vehicle to the rear of the second, the empty space between the two must be resolved. To do this, the interrogation frequency must be at least as high as the reciprocal of the time period for the empty space. Thus, the frequency becomes a function of the desired accuracy of counting a distribution of arrivals in which a specified empty space time period (or less) has a certain probability of occurrence. The following example indicates the manner in which this frequency can be computed.

Suppose the time headway between vehicles moving at 40 mph is 2 sec. Then, in a uniform distribution of headways between vehicles in separate lanes there will be 1 percent of the distribution in each 0.02 sec of headway. Therefore, to arrive at a counting accuracy of 99 percent at a confidence level of 1.0, the interrogation frequency would have to be 50 cps or higher.

Some difficulty that might be experienced with voltage level (or phase shift) discrimination, as shown in Figure B-1, might be eliminated if the count is made on the differential of the voltage (or phase change) with time. Certainly it would seem reasonable to expect that a vehicle entering and leaving a loop of short length compared to that of the vehicle would cause a fairly sharp rise in signal, followed by some shallow undulation and then a sharp drop. If, then, the differentiated voltage rate is plotted, the signature would appear as in Figure B-3. Some minimum value of de/dt is established as a threshold.

The plot of de/dt vs t shows that the leading edge of the two signals that were used as signatures would appear to the detector as coming from one vehicle. However, the

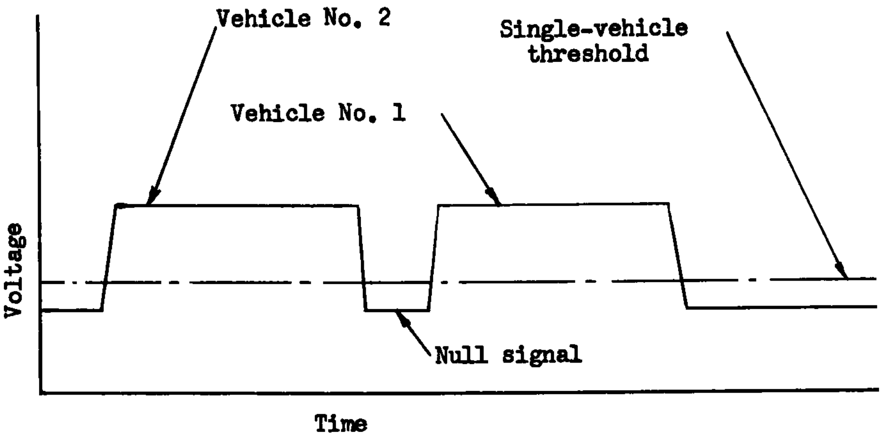


Figure B-2. Single loop, two traffic lanes.

number of times this might occur in a random distribution of traffic in two lanes would be very small. Not having any signatures to examine, the time period over which *de/dt* might exceed some arbitrary value cannot be computed, although it is certain to be quite short. For instance, if full voltage rise took place as the first 2 ft of a vehicle entered the loop at 40 mph, the time period would be 0.033 sec, calling for an interrogation frequency of 30 cps just to "see" *de/dt*. If the computer was to be able to discriminate between successive vehicles at close proximity in adjacent lanes, the interrogation frequency should be about double this value.

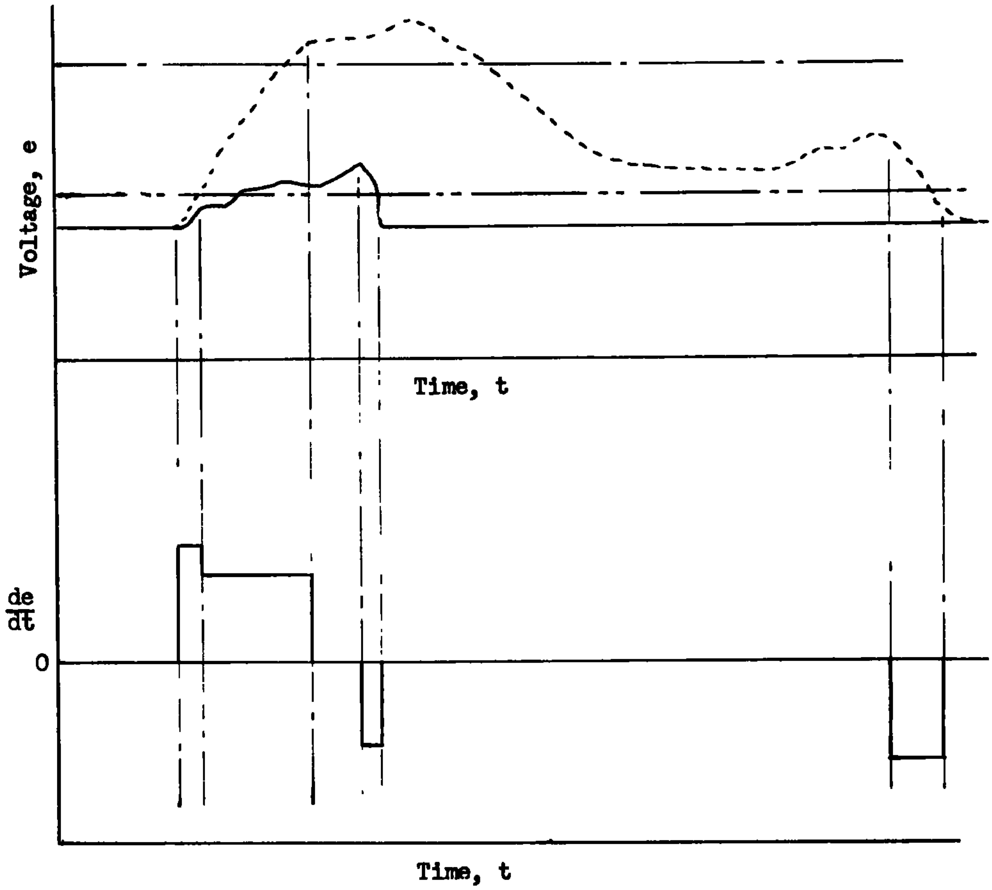


Figure B-3. Analog trace and time derivative.

From the foregoing discussion it is surmised that a 60-cps interrogation rate may serve as an initial value for design purposes. Thus, the total amount of information that can come into the computer during a sampling period will be the frequency times the sampling period times the number of detectors, or $60 \times 1 \times 464 = 27{,}840$ bits per second.

DATA LOGGING

The data sampling period is assumed to be 1 sec. This time period is considered because it would be close to the frequency of arrivals or departures at saturated flow on a typical small city street. It is estimated that such a city would have main streets on which there would be two active lanes on the approaches and that saturation could approach 1,800 vph.

Thus, if the detector interrogation rate is 60 cps, logging of time occupancy may require storage of as many as 60 "vehicle presence" indications. This can be accommodated by 6 binary bits. Since practically any digital computer of reasonable size will have a core storage of 12 or more bits per word, 464 words in the core will hold the raw detector inputs. This assumes that the register will store in binary form.

If the detector can discriminate at three levels, the input logic becomes more complicated. Presence may be determined on 2 lanes. In the simplest case, the detector would put out the information "0", "1", or "2." Thus, the computer would have to accept data through a translator so that it could add 1 or 2 counts when required. Doubling the storage capacity for the channel would mean that 7 binary bits in each word would be used. Computed occupancy would be average for the two lanes, because no lane discrimination is afforded by a single loop detector covering two lanes.

Determination of volume counts on a single lane can be made through a logic that requires that two successive "0's" be followed immediately by two successive "1's" to register a vehicle count. This is consistent with the radar detection philosophy of "two successive blips on two successive scans," and should suffice for traffic surveillance purposes.

A similar logic can be used on a multilane detector. The input will be held over a period of four interrogations, as previously, the logic simply checking that the second pair is at higher level than the first pair. When this condition is satisfied a volume count is added into the register. The volume register need only provide for 1 count per second per lane.

Discussion of the *de/dt* method for counting arrivals and departures is not treated at this time, nor is the complication caused by analog signature for multilane application of the loop detector.

At least one other piece of information would come in from each intersection. This would be the current status of the traffic signal; red, green, or amber, and the identifying direction. Other information would be forthcoming for signals operated on cycles with more than two phases. A logic can be programmed into the computer to determine the elapsed time for the signal phase on each approach from information on just one approach, if a simple cycle and fixed amber phase is used.

Communication of signal condition can be by tone transmission, step voltages, or separate lines. An encoder will be required at the signal end of the line for tone transmission and/or a decoder at the computer for either stepped voltage or tone transmission. Decisions on which arrangement to use would depend on relative cost and reliability.

Thus far, each detector will require two words of core storage, one for occupancy and the other for volume. In addition, an undetermined core capacity will be required to add occupancy bits into the register and to conduct the logic for adding volume counts into its register.

A typical fixed-time cycle controller with provisions for manual control can be operated by the computer control system by two commands. One will be the command to convert from fixed-time cycle to manual (energize a holding relay; failure reverts to fixed cycle) and the other is the command to change phase (a power pulse long enough to energize the solenoid that drives the phase change cam). Three signals (or their absence) will inform the computer of the signal state, assuming a simple four-approach signal light with no special phases. (Such a signal operation is discussed more fully in a succeeding section.) The information that will have to be stored concerning each signal is the state and the last time the state changed. There are four states—red-green, red-amber, green-red, and amber-red—but only the times at which red changes to green, on either set of faces, is essential.

LOCAL CONTROLLER OPERATION

Much or most of the actuation of the traffic signals can be done by fixed-time cycle controllers equipped with a manual control override feature.

In a typical controller an induction motor drives a dial through a reduction gear so that the dial makes one revolution for a full signal cycle. Cycle length is determined by gear selection. Relative phase lengths for the crossing approaches are determined by the relative arcs between pin positions on the dial. As each of the pins reaches an index point it energizes a solenoid that drives a cam through a ratchet. Appropriate rises or falls on the cam open or close the light circuits when the new index position has been reached.

It can be seen that the computer can take over the operation of the local controller by only two command lines. One command signal would change the operation of the controller from fixed-time cycle to computer control, or the reverse. The other command signal would take the place of the phase change pulse from the manual control.

By interrogating three of the actual bulb circuits to the signal faces, the computer can be informed unambiguously as to the state of the signal. For instance, if the red and green circuits on one face and the green circuit on the other face are sampled for voltage or no voltage, the computer logic shown in Figure B-4 will define the state of the signal.

Other local controllers that might be built on the same principle (that is, a timing dial and solenoid-operated

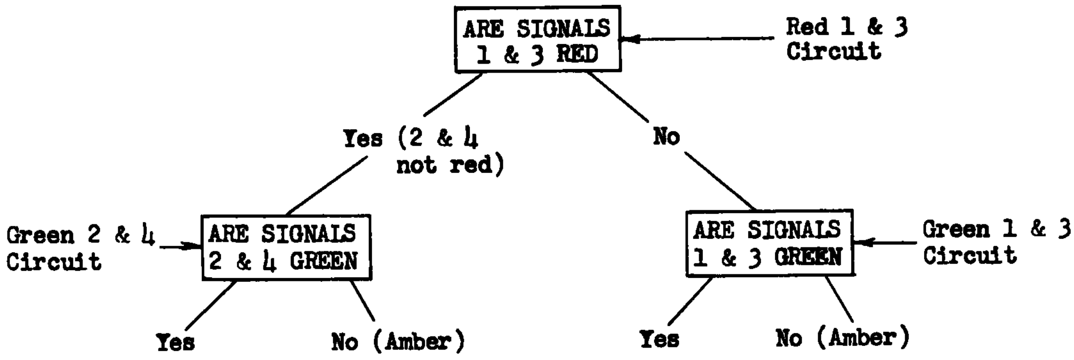


Figure B-4.

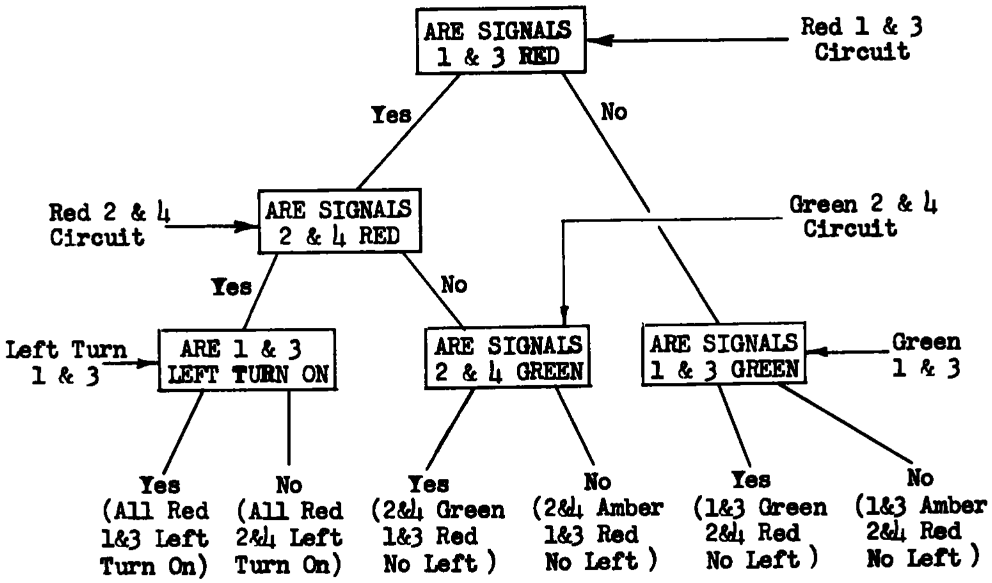


Figure B-5.

phasing cam) can be slaved to the computer in the same way. As the complexity of the phasing increases, the amount of information on the state of the signal will increase, which will require additional lines or multiplexing of tones coming into the storage equipment. For any single case, the best way to transmit the information will be a function of the number of bits coming in and the distance over which they must be transmitted.

For instance, suppose a usual four-way intersection signal includes separate left-turn phases. Information on the state of such a signal would require that two more "bits" than for the ordinary signal be transmitted, one for red on all four faces and one to identify which pair of approaches is showing the "turn left" signal. The logic might be as shown in Figure B-5. "Leading" or "lagging" green signals might be handled in much the same manner, the differences being in the lobing of the phase change cam.

FIXED AND INITIAL INPUT DATA

For each intersection the following items should represent the maximum number of parameters for which values must be placed in permanent storage in the computer. Some are geometric (see sketch on next page):

s_1 = distance from mid-block detector to stop line, approach leg 1;

d_1 = distance to previous intersection, approach leg 1;

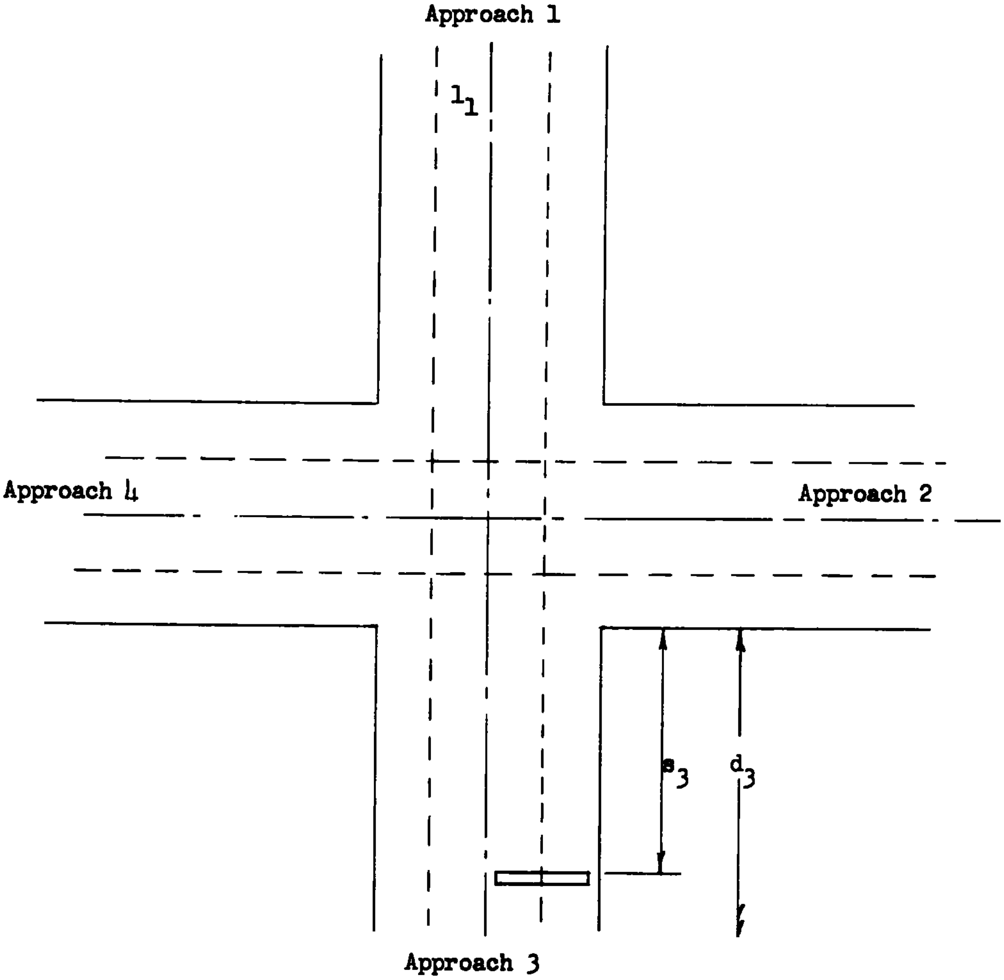
l_1 = number of lanes, approach leg 1 (this may be changed during the day and week if parking rules affect the number of traffic lanes);

and others reflect traffic and control parameters:

$\phi_{G1_{min}}$ = minimum green time, approach leg 1;

$\phi_{R1_{max}}$ = maximum red time, approach leg 1;

ϕ_{cycle} = maximum cycle length, regardless of traffic demand (to check operation of controller in the absence of demand signal);



l_1, r_1 = percentage of left (right) turns from approach leg 1 (this figure may be varied with time of day and day of week, if not determined by surveillance);

v_1 = average approach speed, volume dependent or computed from occupancy data;

$\phi_{A_{min}}$ = minimum clearance time to reflect pedestrian requirements and/or braking conditions and local speed limit.

Departures from these inputs will be required or desired, depending on the nature of specific intersections. For instance, the use of an average of four detectors per signalized intersection is based on the need for up to six at some locations and correspondingly less at others.

PEDESTRIAN CONTROL

In the most general case of traffic control at an intersection, if there is no prohibition to the turning movements of vehicles, the only way in which passage can be guaranteed to pedestrians is by the use of an all-red period long enough to allow people to walk the longest distance from curb to curb, which is the diagonal. The pedestrian problem is of sufficient importance in large metropolitan areas that such special considerations must be included. It is not likely, however, that either traffic density or pedestrian volumes will be so great in the smaller cities as to warrant the use of control doctrines that take into account pedestrian requirements at intersections.

Actually, the only thing that will interfere with pedestrian crossings is a large percentage of turning movements. If turning is moderate, people can make the crossings with the light. Although there may be some slowing of the traffic making turns, assuming that the pedestrian has the right of way, the average flow rates for right turns can be observed and set into the computer control doctrine as a performance number. No account is taken of the effect of left turns, because it is assumed that these would be forbidden at times and locations of heavy traffic.

It is possible that a pedestrian crossing may be required at some mid-block location. The treatment of such a situation would be to include a manually triggered signal in the system. If the crossing was located in a heavily traveled street, it would appear to be appropriate to tie the signal into the rest of the control system. For instance, if the signals along the street were operating in a progressive mode, the pedestrian's operation of the signal would be restricted to signifying a need for passage. The computer would include the changing of the crosswalk signal so that it synchronized as well as possible with the expected gap in the traffic stream. At times when the signal system is not computer controlled the pedestrian signal could be either uncoordinated or tied in with a local progressive controller.

Returning to the idea of an exclusive pedestrian movement phase, this can be accomplished by an extended all-red period. Informing the computer of this state will require four tones from the controller box. If such a phase is built into a controller that provides exclusive left-turn movements, no additional information is required, but only a change in the logic.

APPENDIX C

INPUT INSTRUMENTATION REQUIREMENTS FOR A REAL–TIME DIGITAL–COMPUTER–CONTROLLED TRAFFIC SIGNAL SYSTEM

To implement the traffic control equations of Appendices D, E, and G in a complex network of street intersections by means of a centralized computer control it is necessary to have some form of information gathering equipment at the critical intersections. This equipment may be divided into two groups; data collection and vehicle control. The traffic state at any intersection must be known at all times. From computations based on this information the computer sends to the signal control units at the intersections control signals intended to optimize traffic flow.

INTERSECTION INSTRUMENTATION

A general four-way intersection is shown in Figure C-1. This represents the maximum amount of detection equipment that would be required at a single installation. Many intersections would be of the T or Y type and some would involve one-way traffic. In these instances less detection equipment would be required. Figure C-1 shows eight vehicle detectors, four located at the stop line of the intersection and four upstream on the approach lanes to the intersection. The detectors are of the induction loop type, which initiates detector contact closure when a vehicle passes over the loop. To minimize the number of detectors necessary, a time-sharing technique would be employed where the two detectors in the intersection would be switched between the four stop-line loops, depending on the phase of the signal controller. When the signal controller was green for the N-S traffic, loops 1 and 3 would be active; when the signal controller became green for the E-W traffic, loops 2 and 4 would be active. The approach detectors would be active at all times, thereby providing continuous data on incoming traffic.

Figure C-2 shows a typical traffic phase pattern and a block diagram of the logic necessary to accomplish the switching function. Loop detector A is connected to either loop 1 or loop 4, depending on the signal controller condition. If the N-S signal is either amber or green a signal passes through OR gate 3 and enables AND gate 1, thereby connecting loop 1 to detector A through OR gate 1. If the E-W signal is either amber or green a signal passes through OR gate 2, which enables AND gate 2 and connects loop 4 with detector A through OR gate 1. A typical suggested implementation using relays as the logic elements is shown in Figure C-3.

The instrumentation for detection of the phase state of the signal controller is shown in Figure C-4. Only three conditions are needed for the complete specification—detection of green condition on the N-S face, green condition on the E-W face, and red condition on the N-S face. Because the signal controller is either on or off in each case, a binary coding scheme may be used. A phase pattern and coding table also are shown in Figure C-4.

INPUT LOGIC AT CONTROL STATION

The data from each intersection must be transmitted to a central control station where the computer is housed. Because all information is of an on-off (binary) nature, a selective signaling system using telephone lines was adopted as a transmission medium. Each channel of information would have a specified signal frequency assigned. At the central site tuned receivers would be used to detect the information. Each telephone line would then carry a composite of tone signals in both directions between intersections and central control.

The information from each intersection is decoded and grouped into a computer word in the following manner:

1. Bits 1-3 contain the three bits denoting the phase state of the signal controller.
2. Bits 4, 6 contain the time-shared stop line intersection vehicle detection information. Vehicles are counted only in the green phase direction.
3. Bits 8, 10, 12, 14 contain the approach vehicle detection data, which are continuous from each intersection.
4. Bits 15-21 contain the permanent binary number identifying each intersection. These seven bits offer assurance that the computer will recognize correctly the data from the individual intersections.

Figure C-5 shows the input word structure.

The number of instrumented intersections totals 116. Data from each are brought continuously to the central control site and the computer control commands are sent

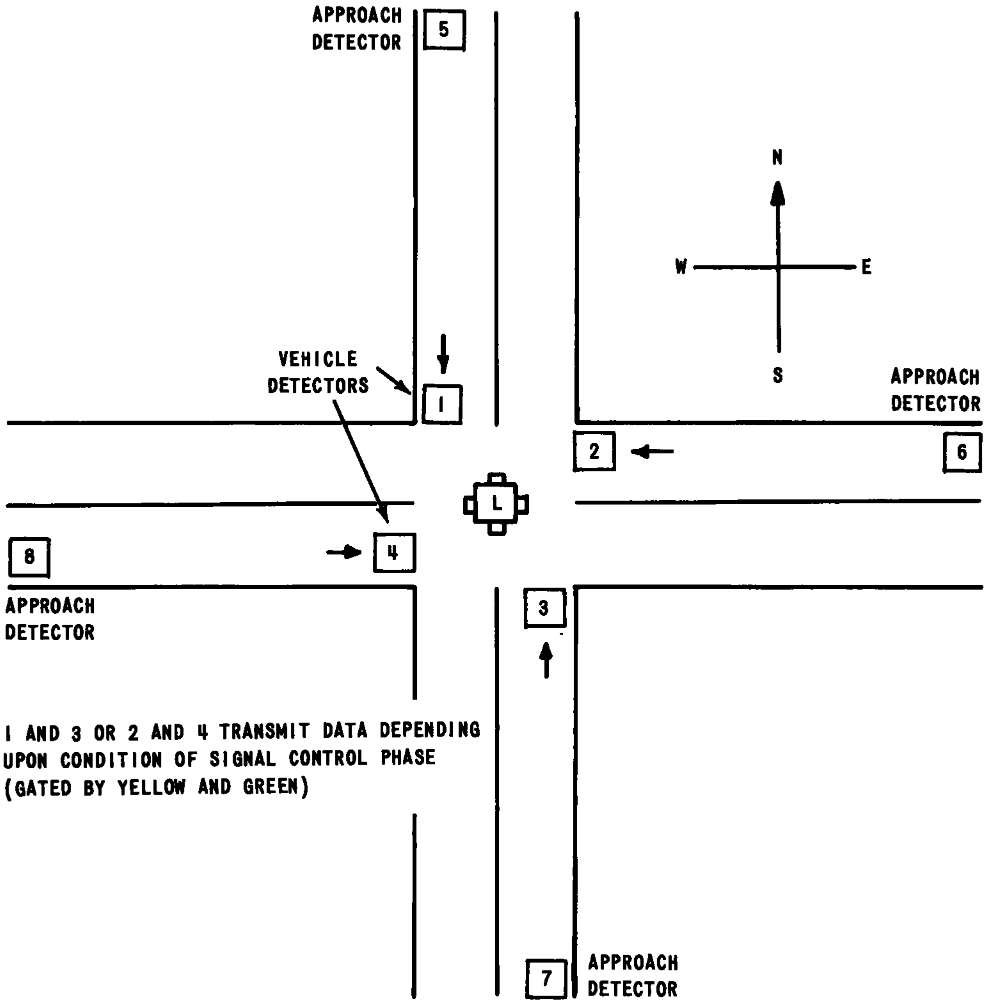


Figure C-1. Intersection configuration.

back to each intersection as required. To get the required accuracy to meet the control system specifications, it was desirable to sample vehicle detectors approximately 60 times per second. Each sampled input data word is then presented to the computer input terminals for approximately 144 μsec (microseconds). This is the multiplexer period; the product of the inverse of the sampling frequency and the number of sampled intersections ($T = 1/60 \times 1/116$). Figure C-6 is a logic diagram of the input circuitry. The basic clock frequency of the chosen computer is 500 kilocycles. This is divided down by a factor of 72 to arrive at the multiplexer gating frequency (6,941 cps), which approximates the product of the sampling frequency and the number of sampled intersections (60×116). This in turn drives a 7-bit counter, which resets after 116 counts. The true and false outputs of each flip-flop in the counter provide the coding signals for 116 multiplexer AND gates. Each of these gates has seven input lines from the counter and one input from a delayed gate pulse to allow the counter lines time to recover from the transient switching conditions. The output from each of these AND gates goes to another group of 21 AND gates associated with the appropriate 21-bit intersection data word. When a particular multiplexer AND gate has the correct input it enables the associated 21 AND gates, and 21 bits of data for that intersection are presented to the computer input lines through multiple input OR gates. These OR gates have 116 inputs and provide the connection between intersection data lines and the computer input lines.

OUTPUT LOGIC AT CENTRAL CONTROL STATION

Figure C-7 shows a logic diagram of the output circuitry. The output control signals from the computer are generated at a rate of one every 2 sec. They consist of a HOLD command and an ACTUATE command. Each consists of one bit in the output word structure. Bits 3-9 are the intersection identification information. There are 116 selector AND gates which have 8 input lines each. Seven of the lines connect to the intersection identification bits in the output word. The eighth input is derived from a computer-generated Read pulse gated by the Actuate command. The output of the selector gate provides one input to 3 AND gates controlling the HOLD and ACTUATE signal circuitry. When it is desired to transfer a signal controller at a particular intersection from fixed-time operation to computer control, the output word presented by the computer will contain a binary 1 in the HOLD bit and the correct 7-bit intersection data. The HOLD bit enables one input to an AND gate driving the Set input to a flip-flop circuit. The Read pulse provides the trigger signal which sets the flip-flop. The flip-flop output is then transmitted to the intersection, transferring signal controller operation to the computer. If it is now desired to change the controller phase state the ACTUATE bit becomes binary 1. The Read pulse now goes to an AND gate triggering a multivibrator, which generates a control pulse of long enough duration to cause an actuation command at the remote signal controller. A duration of from 300 to 500 milliseconds is considered sufficient. This pulse is transmitted to the intersection and actuates a stepping circuit in the signal controller which changes the phase. To return a signal controller to fixed-time operation the HOLD bit becomes binary 0. This enables an AND gate connected to the Reset input to the hold circuitry flip-flop. The next Read pulse then resets the flip-flop and the controller is released from computer control. The computer will evaluate each of the 116 subnetworks in sequence. Actually, the decision of the phase change will be dependent on the value of the objective function (i.e., relative delay) at any given time, determined by the current and predicted traffic flow in the system (Appendix F).

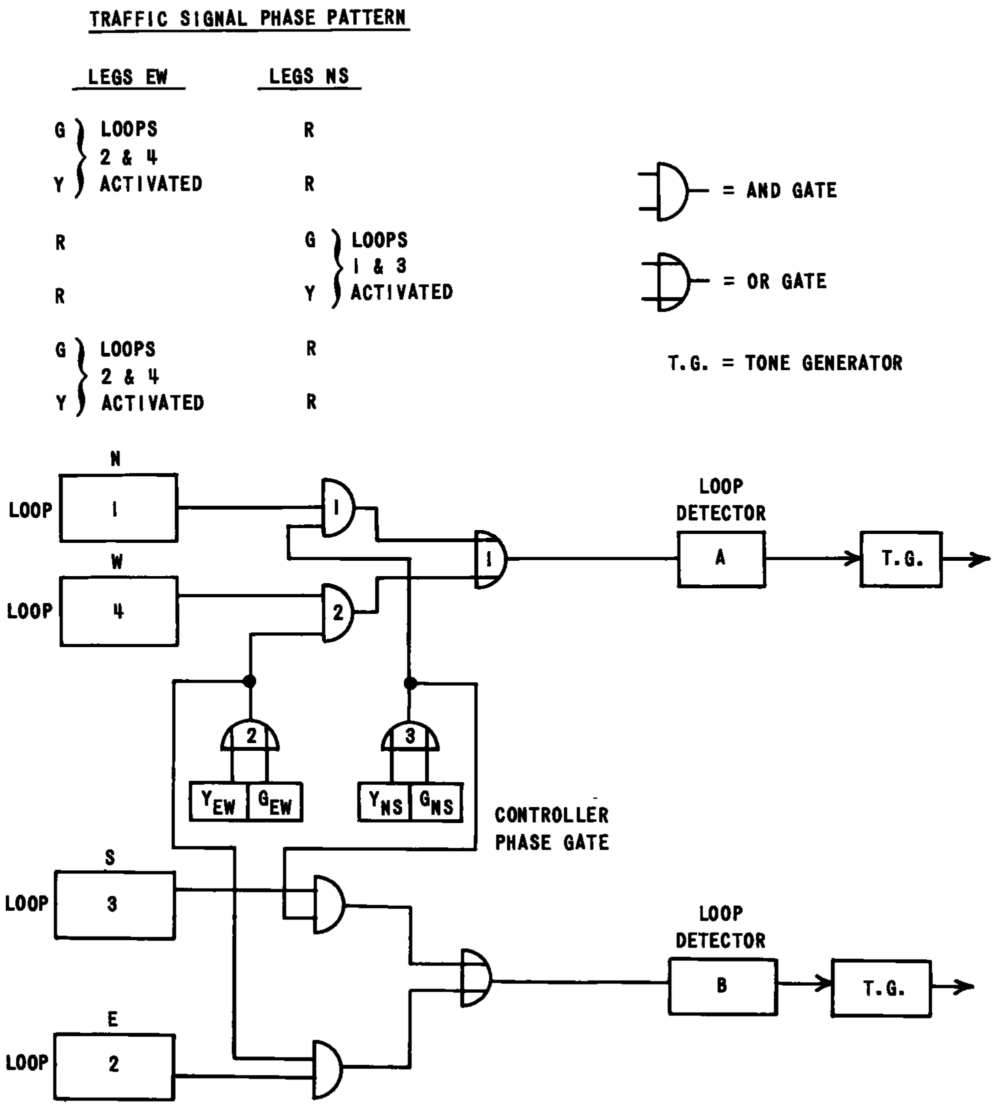


Figure C-2. Time-sharing vehicle pickups using signal controller phase as gate.

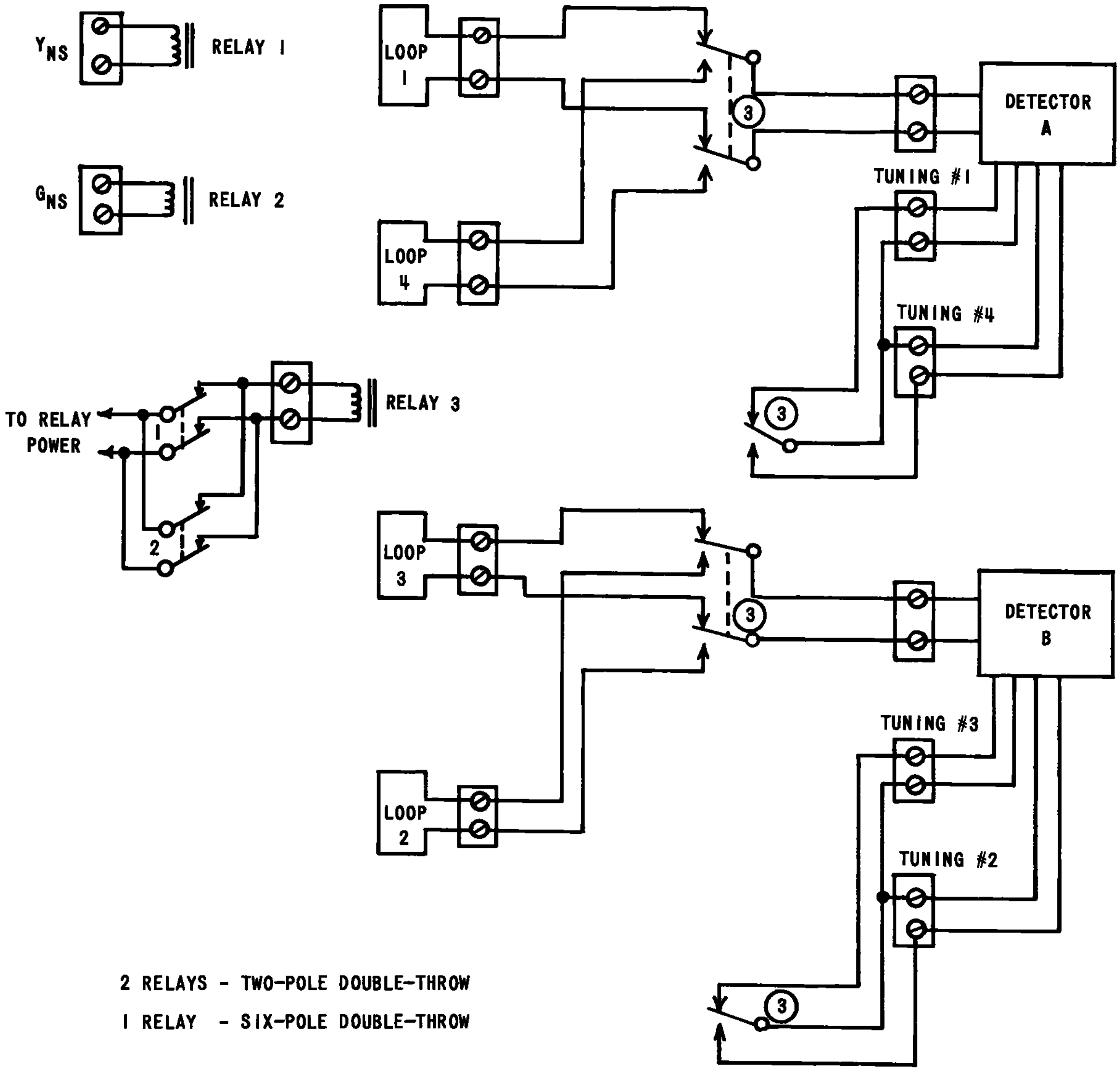


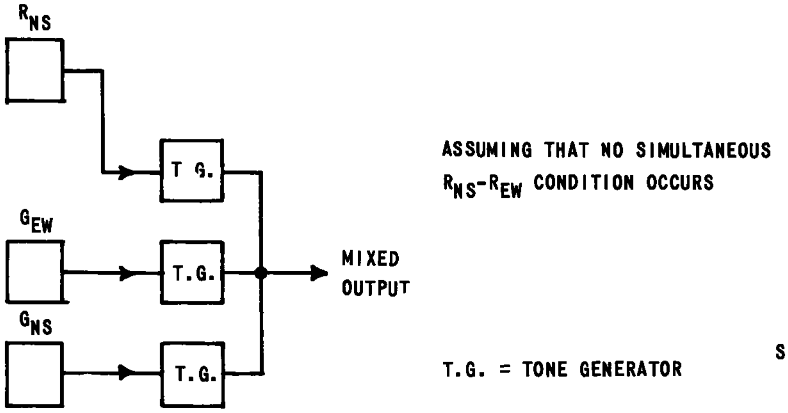
Figure C-3. Suggested hardware for time-sharing vehicle detector units at intersection.

SYSTEM INSTRUMENTATION COST

An investigation was made with reference to the sources of supply for the instrumentation required at each intersection and at the central site. Based on the mode of operation desired, a typical set of equipment cost figures was compiled for the complete installation. The breakdown was as follows:

Intersection equipment, @ $1,575	$182,700
Input logic interface at central site	26,530
Output logic interface at central site	19,740
Telephone line leasing	2,000/mo.
Total	$228,970 plus $2,000/mo.

These figures are approximate, as telephone leasing rates will depend on the community selected. Not included are the maintenance and construction costs associated with the equipment's installation and upkeep.



BINARY CODE R_{NS}	G_{NS}	G_{EW}	OCTAL EQUIVALENT OF BINARY CODE
1	0	1	5
1	0	0	1
0	1	0	2
0	0	0	0

SIGNAL CONTROLLER PHASE PATTERN

NS	EW		
R	G	1 CYCLE	5
R	Y		1
G	R		2
Y	R		0
R	G		5
R	Y		1
G	R		2
Y	R		0

Figure C-4. Monitor for phase condition of signal controller.

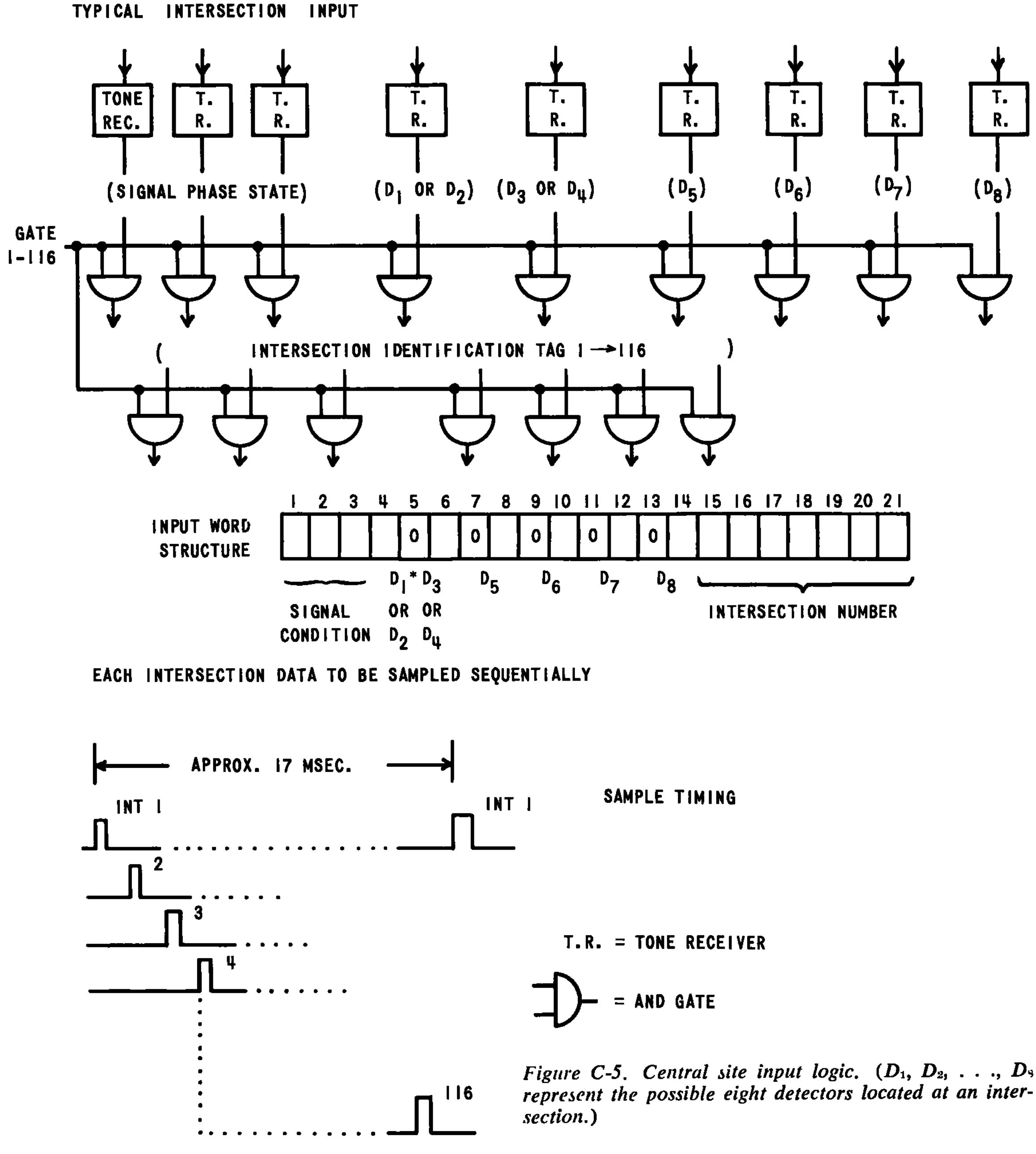


Figure C-5. Central site input logic. (D_1, D_2, . . ., D_8 represent the possible eight detectors located at an intersection.)

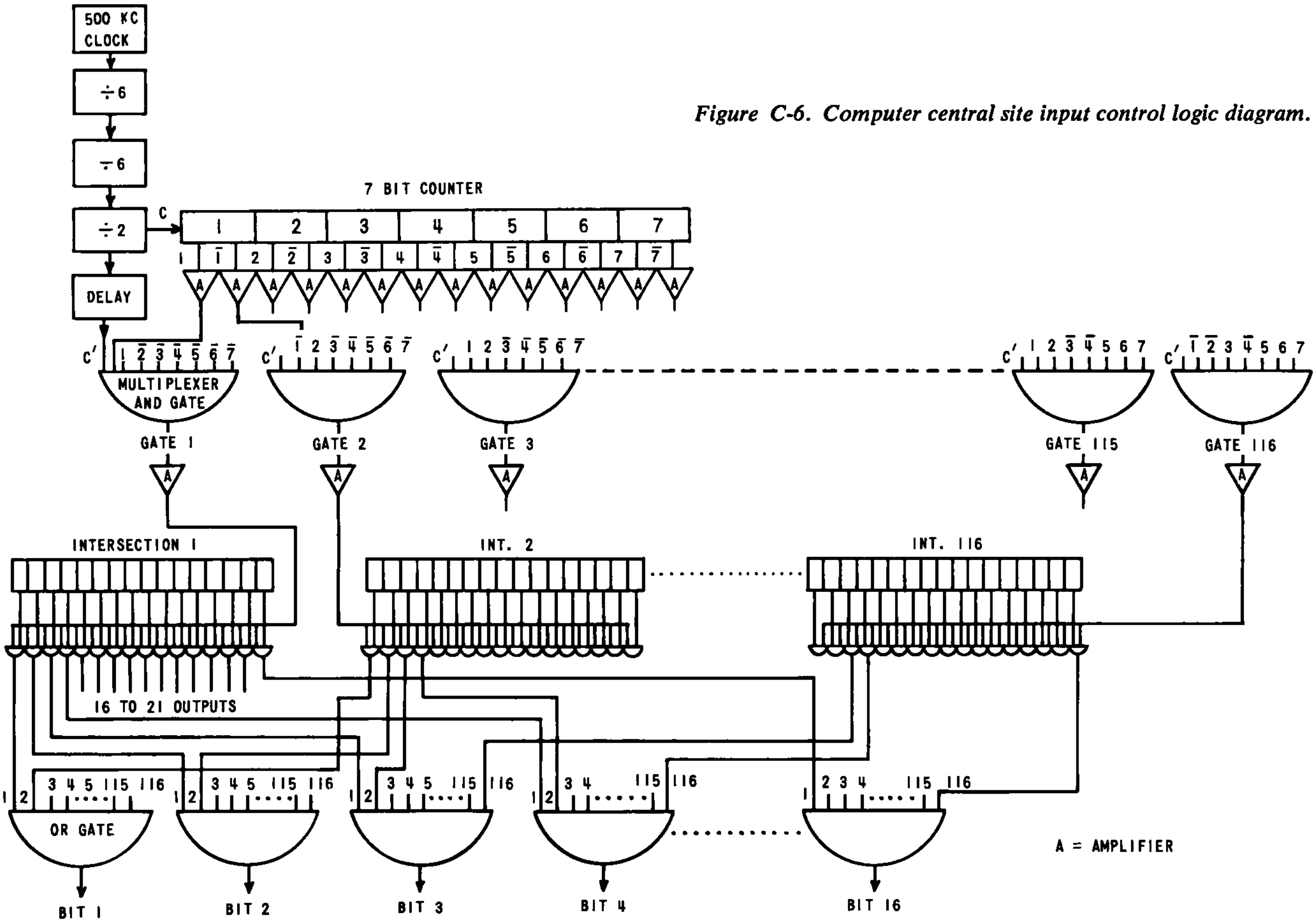


Figure C-6. Computer central site input control logic diagram.

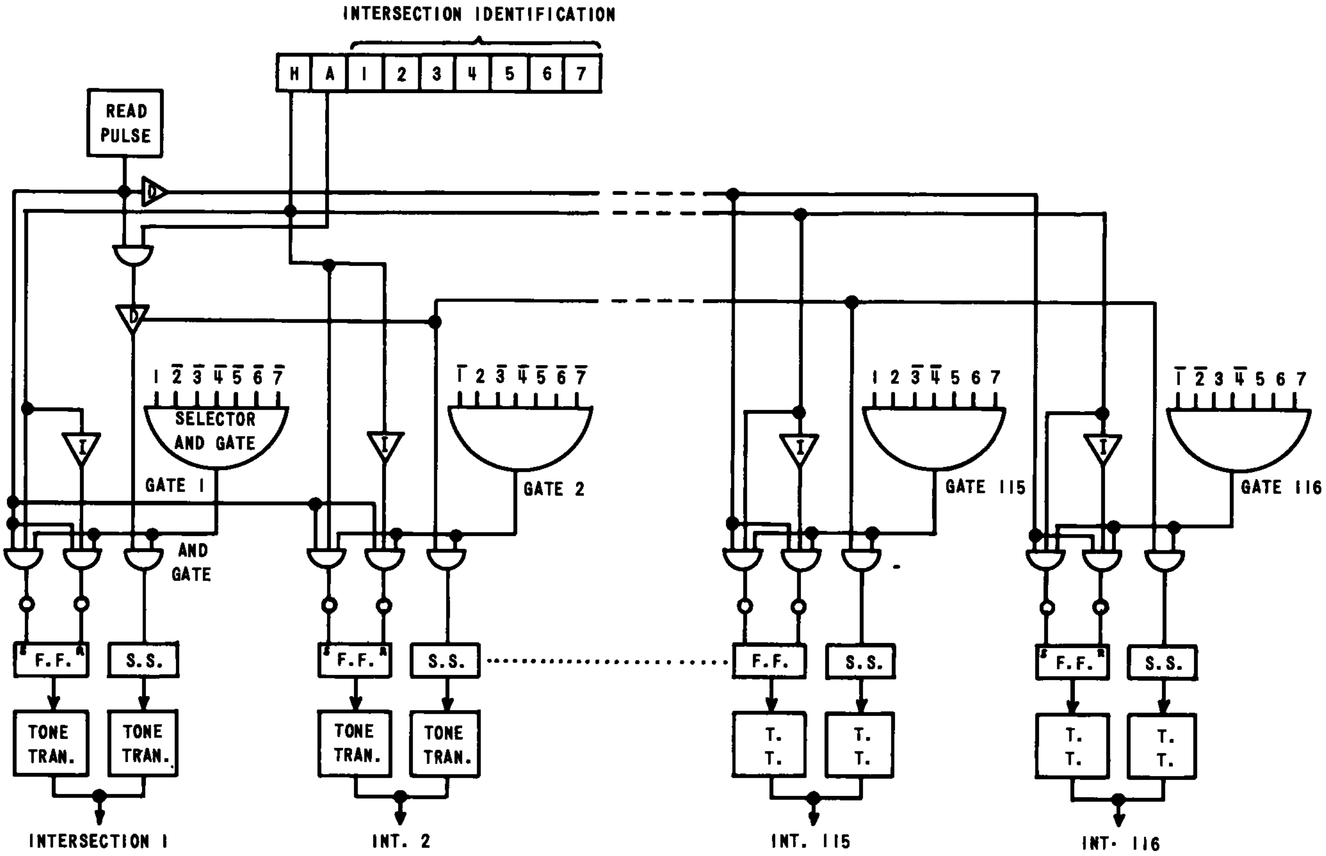


Figure C-7. Computer site output control logic diagram.

APPENDIX D

DERIVATION OF A MINIMUM-DELAY FUNCTION FOR AN ISOLATED INTERSECTION WITH SIMPLIFYING ASSUMPTIONS

It has been estimated that more than 75 percent of vehicular delay in urban areas can be attributed to intersections in general. However, due to the geometric grid network and the associated volumes of vehicles, such areas contain a high concentration of signalized intersections. Without changing the physical characteristics of the network, the efficient operation of these signals then becomes a major factor in the reduction of the avoidable delay.

The primary function of a traffic signal at an intersection is to regulate capacity (assign right-of-way). Selection of proper cycle lengths and splits for these signals is at times critical and in general necessary if service to the motorist is to be optimized. The term "optimization" in this context is used to mean the minimization of the motorists' travel time or delay (i.e., time necessary to travel a given route in excess of that required for "free" vehicles). The existence of such a delay, as defined, is inevitable with the current static network (e.g., physical characteristics, single level intersection).

Numerous attempts have been made to solve the signalized intersection problem. Hence, there are currently many types of specialized traffic signal equipment and systems whose purpose is to "expedite" traffic through urban intersections. Although many such systems have produced good results, others have at times systematically aggravated rather than alleviated a specific traffic problem. In any case, specialized systems implement discrete control philosophies which are artifices employed to obtain the minimal-delay results and are too limited in adjusting to the rapid fluctuations of urban traffic. To obtain a more desirable system having the sensitivity and flexibility to react to the uncertainty and variability of traffic flow that occurs within the time intervals of practical signal phase lengths, it is proposed that a computer directly implement the basic delay criterion. Logically, the control decisions should result from the continual sampling of the traffic flow characteristics and with such speed as to gather the pertinent data, calculate the decision, and actuate the controls in real time. To process the large quantities of data within the required time intervals, the system will be synthesized employing a high-speed large-scale digital computer. Information describing the traffic flow and signals displayed will be fed into the computer from strategically located detectors. The detectors may measure the number of vehicles, speed and/or presence to convey the "state" of the network to the computer. The computer, working in real time through the programmed equations implementing the minimum-delay criterion, will compute the best set of signals at a given time for the entire network.

Hence, the proposed system may be viewed as a servomechanism (Fig. D-1) in which the dynamic network of vehicles, pedestrians, weather, etc., is monitored (within the limitations imposed by the detectors employed) with the computer operating on the dynamic and static information and the resultant signal display relating the decisions to the motorists and pedestrians. The advantages and limitations of any such system (in vehicle-seconds of delay saved) is difficult to evaluate analytically. Specifically, it is suggested that the relative merits (when compared to less costly alternatives) of a particular system, be evaluated by computer simulation techniques (accounting for the limitations inherent in any model). The ultimate test, of course, is in its application to the actual city being studied.

Another approach in the synthesis of such a system is to program the computer to simulate discrete traffic control techniques as currently employed by individual specialized systems (e.g., progressive movement, linked volume-density, etc.) With this method, the computer would necessarily be capable of assigning various control subroutines to different intersections as dictated by traffic flow (sensors), time of day (clock), or the decisions of a human operator. Heuristically it appears that the more general approach proposed here (i.e., the implementation of the basic criteria in the form of a set of equations), would be the most flexible of the two methods, and in general would have the smaller reaction time to changes in traffic demands.

Currently, researchers are experimenting with each of these general philosophies (*3, 4*). It is hoped that they will be able to contribute to the evaluation of the relative advantages of the two approaches.

DERIVATION WITH ASSUMPTIONS

The starting point is observation of signalized intersection A, which may be assumed independent of any other intersection (i.e., isolated). With the proposed control system the digital computer is periodically fed information from remote sensors that serve to sample the dynamic state of the intersection. Based on the minimum-delay criterion, the system uses all pertinent information (e.g., updated detector information, and information from the computer's permanent memory) to make a decision as to which signal display would be optimum. As a result of these computer calculations, local signals are controlled by actuating relays at the particular intersection through the system's communications network. For the present example it is assumed that only signals indicating stop (red), go (green) and clearance (amber) are available for display to complementary traffic (e.g., north-south or east-west). Hence, exclusive pedestrian phases, turning arrows, etc., are not included.

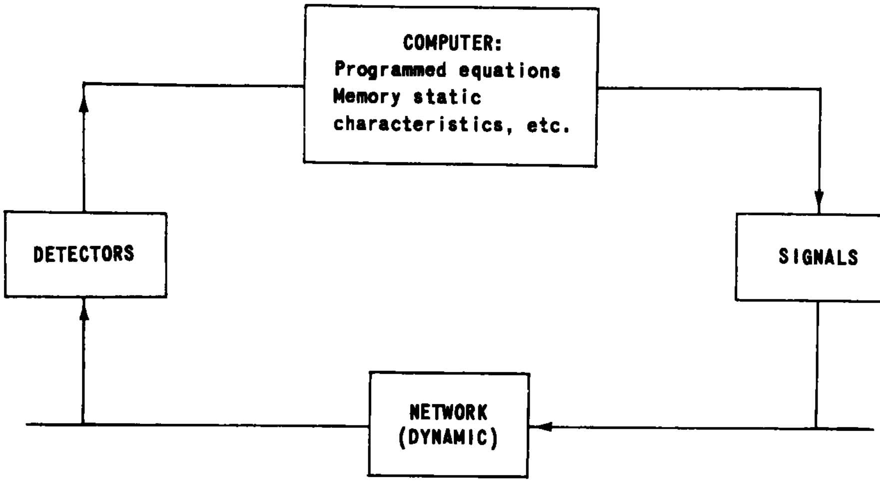


Figure D-1. Traffic control loop logic flow diagram.

The traffic signal's phase length, although flexible and under the control of the computer, will be bounded by practical limits for minimum green time, $\phi_{G|E,W_{min}}$, and maximum red time, $\phi_{R|N,S_{max}}$. Within the limits * ${}_A[\phi_{G|E,W_{min}} \leq \phi_{tG|E,W} \leq \phi_{G|E,W_{max}}]$ the computer is free to decide on the optimum cycle length and split, after considering the traffic conditions and intersection geometry. It should be noted that $\phi_{tG|E,W}$ represents the elapsed green time of the east-west signals and $\phi_{G|E,W_{min}}$ represents an interval of time chosen to be as small as practical, taking into account the drivers' perception, reaction (inertia), and travel time for the particular intersection characteristics. The upper limit, $\phi_{G|E,W_{max}}$, is chosen to be as large as practical in order to allow the computer maximum flexibility. This quantity, $\phi_{G_{max}}$, is a function of the human factor dictated by the maximum time an individual may be required to "see red" without suffering potentially serious irritations (i.e., uncertainty as to whether or not the signal is operating, frustration, anxiety).

It is proposed that the computer interrogate the intersection (e.g., detectors, signal display, elapsed phase-time) at time intervals of Δt. This interval should be small enough that calculated decisions are implemented in what may be considered real time, yet large enough that a modern high-speed computer will be able to update and/or adjust all other signals in the system during the same time interval. The derived equations must necessarily calculate the relative delay which would be suffered if the signals are changed immediately, or in Δt, or in integral multiples of Δt.

Now consider the relative aggregate vehicle delay time saved to east-west traffic, ${}_AS_{T|E,W}$, which would result by extending the $\phi_{G|E,W}$ phase by Δt as compared to the immediate termination of $\phi_{G|E,W}$. (The vehicle delay time may be expressed in vehicle-seconds.) This term may be considered in two separate parts: the first term, S_D, is a result of the delay saved due to the departing vehicles during the proposed extension; the second term, S_q, is the resultant savings that accrue to the vehicles forming in a queue behind those departed and are served by the intersection at an earlier time, once the signals have again changed.

* ${}_A[\ \ ]$ = with respect to intersection A

Hence, the total (relative) delay saved for intersection A is

$$ {}_AS_{T|E,W} = {}_A[S_D + S_q]_{E,W} \qquad \text{(D-1)} $$

The S_D term expresses, in functional form, the (estimated) delay saved by those vehicles serviced (from the two legs) during Δt, which otherwise would have to suffer the amber, red, and lost time resulting from an immediate change in phase. Hence,

$$ {}_AS_D = {}_A(\tilde{d}_E\Delta t + \tilde{d}_W\Delta t)\ {}_A(\tilde{\phi}_{R|E,W} + \tilde{\phi}'_{A|E,W} + \tilde{l}_{E,W}) \qquad \text{(D-2)} $$

in which

$\tilde{d}_E, \tilde{d}_W$ = estimated departure rates, in vehicles per second, from the east and west legs, respectively, during the proposed extension, Δt;

$\tilde{l}_{E,W}$ = estimated lost time attributed to initial acceleration period from a stopped position (represents the aggregate additional time experienced by the first few vehicles);

$\tilde{\phi}_{R|E,W}$ = estimated duration of next red phase, in seconds;

$\tilde{\phi}'_{A|E,W}$ = estimated portion of amber phase that may be considered an extension of $\phi_{R|E,W}$; and

$\sim$ = an estimate of.

It should be noted that no consideration has been given to multiple lanes (i.e., vehicle distribution). For simplicity, it is assumed that vehicles heading in a particular direction have only a single lane in which to travel. This development is extended to include the more general case in Appendix E.

The S_q term of Eq. D-1 expresses, in vehicle-seconds, the delay saved which accrues to those vehicles served earlier in time as a result of the previous departures. The term must therefore calculate the total number of arrivals during the next phase, $\phi_{R|E,W}$, plus those remaining in queue, and multiply this sum by the time the vehicles are nearer to service. Hence,

$${}_{A}S_{qE,W} = {}_{A}(S_{qE} + S_{qW}) \tag{D-3}$$

where

$${}_{A}S_{qE} = \left[\left(\sum_{i=1}^{L} a_{iE}\Delta t_i\right) + N_{rE}\right] \frac{\tilde{d}_E \Delta t}{\tilde{d}_{L^+E}} \tag{D-4}$$

$${}_{A}S_{qW} = \left[\left(\sum_{i=1}^{L} a_{iW}\Delta t_i\right) + N_{rW}\right] \frac{\tilde{d}_W \Delta t}{\tilde{d}_{L^+W}} \tag{D-5}$$

a_{iE}, a_{iW} = arrival rates, in vehicles per second, on the east and west legs, respectively, the numerical values, measured and/or estimated for the i th interval, being constant for Δt_i;

$\tilde{d}_E, \tilde{d}_W$ = estimated departure rates, in vehicles per second, from the east and west legs of the intersection during the proposed extension, Δt;

$\tilde{d}_{L^+E}, \tilde{d}_{L^+W}$ = estimated departure rates, in vehicles per second, from the east and west legs during the initial portion of the next phase, $\phi_{G|E,W}$;

N_{rE}, N_{rW} = residual queues on east and west legs, in number of vehicles;

S_{qE}, S_{qW} = delay saved on the east and west legs as a result of the vehicles being served nearer to the start of the next green phase, $\phi_{G|E,W}$;

$$L = \frac{(\tilde{\phi}_{R|E,W} + \tilde{\phi}'_{A|E,W} + \tilde{l}_{E,W})}{\Delta t} \tag{D-6}$$

calculated to the nearest whole number.

The total saving, in vehicle-seconds, may now be written as

$$S_{T|E,W} = S_{D|E,W} + S_{q|E,W} \tag{D-7}$$

or

$$\begin{aligned} S_{T|E,W} = {} & [\tilde{d}_E \Delta t + \tilde{d}_W \Delta t)(\tilde{\phi}_{R|E,W} + \tilde{\phi}'_{A|E,W} + \tilde{l}_{E,W})] \\ & + \left[\left(\sum_{i=1}^{L} a_{iE}\Delta t_i\right) + N_{rE}\right] \frac{\tilde{d}_E \Delta t}{\tilde{d}_{L^+E}} \\ & + \left[\left(\sum_{i=1}^{L} a_{iW}\Delta t_i\right) + N_{rW}\right] \frac{\tilde{d}_W \Delta t}{\tilde{d}_{L^+W}} \end{aligned} \tag{D-8}$$

The next calculation is to determine the delay, D_{TN} and D_{TS}, caused to the vehicles on the north and south legs, respectively, by deferring the start of phase $\phi_{G|N,S}$ for a period Δt. It is assumed that this delay function, D_T, as applied to any leg may be represented by three separate terms—the delay, D_{dec}, caused by the deceleration of vehicles to join the end of the queue; the delay, D_q, experienced in the queue waiting to be serviced by the intersection; and D_{acc} attributed to the time necessary for those vehicles stopped in queue to accelerate to "free" speed. However, since the previously defined term "delay" was taken to mean the additional travel time above that normally incurred, another factor D_{norm} must be included, representing the minimum number of vehicle-seconds required to traverse the distance corresponding to the application of the delay equation. Hence,

$$D_{TS} = D_{dec\ S} + D_{qS} + D_{acc\ S} - D_{norm\ S} \tag{D-9}$$

Eq. D-9 expresses the delay experienced by vehicles on the south leg due to the additional time necessary above that normally required to travel the same route unimpeded.

The duration of the queue on the south leg, $k_S \Delta t$, as a result of changing the phase to $\phi_{G|N,S}$ immediately, may be found by solving Eq. D-10. The number of vehicles in queue at the start of the new phase on the south leg is given by N_{IS}, and a_{iS} and d_{iS} are the arrival and departure rates, respectively, for the interval Δt_i; that is,

$$N_{IS} + \sum_{i=1}^{k_S} a_{iS}\Delta t_i - \sum_{i=1+l_S}^{k_S} d'_{iS}\Delta t_i \leq 0 \tag{D-10}$$

where k_S is the smallest integer to satisfy the inequality. Note that the departure rate d'_{iS} (e.g., saturation flow) occurs only after an elapsed time of l_S representing the assumption that the time lost in the starting transient may be represented by a period, l_S, of zero departure rate, instead of in actuality being distributed among the first few departing vehicles.

If, however, the start of the $\phi_{G|N,S}$ phase is postponed by a period Δt, the depletion of the new queue on the south leg will occur at time $k'_S \Delta t$. Where k' is the smallest integer to satisfy

$$N_{IS} + \sum_{i=1}^{k'_S} a_{iS}\Delta t_i - \sum_{i=1+l_S+\Delta t}^{k'_S} d_{iS}\Delta t_i \leq 0 \tag{D-11}$$

Similar equations exist for the extended queue duration for the north leg. Hence, the additional time (relative to the alternatives) of the existence of the south queue Δt_{qS} is given by

$$\Delta t_{qS} = k'_S \Delta t - k_S \Delta t \tag{D-12a}$$

or

$$\Delta t_{qS} = \Delta t (k'_S - k_S) \tag{D-12b}$$

and the similar equations for the north leg yield

$$\Delta t_{qN} = \Delta t\ (k'_N - k_N) \tag{D-13}$$

Although all vehicles required to stop at an intersection will suffer a delay (when compared to a free-flow condition), only those vehicles impeded by the extension Δt_q will be considered in the relative delay calculations of the D_{dec} and D_{acc} terms. If it is assumed that all vehicles have similar deceleration and acceleration characteristics, both the aforementioned delay terms will be proportional to the number of additional vehicles stopped. With the assumed constants of proportionality of K_1 and K_2 for D_{dec} and D_{acc}, respectively,

$$D_{dec\ S} = K_1 \sum_{i=k_S}^{k'_S} a_{iS}\Delta t_i \tag{D-14}$$

$$D_{acc\ S} = K_2 \sum_{i=k_S}^{k'_S} a_{iS}\Delta t_i \tag{D-15}$$

where the summation of the arrivals, a_{iS}, between the limits k_S and k'_S yields the additional stopped vehicles on the south leg.

Note that it has been emphasized throughout this development that for the purposes of control decisions the programmed equations should be relative in order to obtain the best solution from the alternatives (i.e., difference in delay occurring in changing the signals immediately or at a time Δt later).

Now examine the term D_q, expressing the difference in vehicle-seconds suffered in the queue for the two alternatives. The number of vehicles in the queue, N, in general may be expressed as a continuous function of time $N(t)$. This representation is physically realistic, as vehicles do not jump out of or into queue, but rather flow (e.g., ⅔ vehicle per second). Hence, Figure D-2 may express the general queue size at an intersection for two specific cases (i.e., the two alternatives), where $N(t)$ represents the length of the queue for the case in which ϕ_G is started at $t = 0$ and $N'(t)$ for the condition when the start of phase ϕ_G is postponed by Δt.

The term D_q may now be expressed mathematically as

$$D_q = \int_0^{k'\Delta t} N'(t)\,dt - \int_0^{k\Delta t} N(t)\,dt \qquad \text{(D-16)}$$

However, to apply Eq. D-16 to the proposed system the integration must be performed numerically. If the assumptions implied in this case by the sampling rate $1/\Delta t$ (i.e., that the predictions and/or measurements of the arrival and departure rates are constant over the period Δt) are included, the exact integral, in numerical form, is obtained in Eq. D-18. The form of Eq. D-17 derives from the following: At any time $j\Delta t$, the number of vehicles in queue, N_{S_j}, will be the initial number, N_I, less the integral of the departure rate minus the rate of arrivals over the period $j\Delta t$. The delay suffered by the vehicles in queue from $j\Delta t$ to $(j+1)\Delta t$ will be the average number of vehicles in the interval multiplied by the interval Δt. Thus, the delay suffered during the interval $j\Delta t$ to $(j+1)\Delta t$ will be

$$\text{delay}_{\,j\Delta t-(j+1)\Delta t} = \left[(N_{jS})_{j\Delta t} - \frac{(a_j - d_j)}{2} \right] \Delta t \qquad \text{(D-17)}$$

Hence, the relative delay equation is

$$D_{qS} = \sum_{j'=1}^{k'_S} \left\{ \left[N_{Ij} + \sum_{i'=1}^{j'} (a_{i'} - d_{i'}) \Delta t_{i'} - (a_{j'} - d_{j'}) \frac{\Delta t_{j'}}{2} \right] \Delta t_{j'} \right\}_S - \sum_{j=1}^{k_S} \left\{ \left[N_{Ij} + \sum_{i=1}^{j} (a_i - d_i) \Delta t_i - (a_j - d_j) \frac{\Delta t_j}{2} \right] \Delta t_j \right\}_S \qquad \text{(D-18)}$$

in which all parameters within $\{\ \}_S$ apply to the south leg, the terms with the unprimed subscripts apply to $N(t)$ of Figure D-2, and those with the primes refer to $N'(t)$. The summation limits start at $i = 1$, hence the unprimed departure rates, d_i and $d_{i'}$, represent the distribution of the aforementioned lost time, l_S, within these terms*.

As an example in applying Eq. D-18, let it be assumed that the alternative to the immediate change to $\phi_{G|E,W}$ is the extension by the amount Δt of the current $\phi_{G|N,S}$ phase. Then $d_{1E} \geqq 0$ and $d_{1E'} = 0$, because by definition the extended $\phi_{R|E,W}$ phase would limit the departure rate from the east leg to zero.

Since the delay equations for the other legs (north, east, and west) are similar, they will not be presented here in detail.

Let it be assumed that a specific physical condition exists, in order to indicate the manner in which a decision may be made employing the previously developed equations. Suppose that the east and west legs are receiving the red phase, $\phi_{R|E,W}$, and the minimum green time, $\phi_{G|N,S_{min}}$ has elapsed. Then the computer's decision to change or extend the current phase by Δt must be based on the minimum-delay criterion, assuming also that the current elapsed green time, $\phi_{tG|N,S}$, is less than $\phi_{G|N,S_{max}}$. The following equation must be calculated by the computer:

$$S_{T|N,S} - D_{TE} - D_{TW} = \theta_{\Delta t} \qquad \text{(D-19)}$$

* The lost time (e.g., time necessary to attain steady-state service) may be accounted for in D_q by allowing the departure rates to reflect the smaller (than saturation flow) values of d until the steady-state condition has been attained.

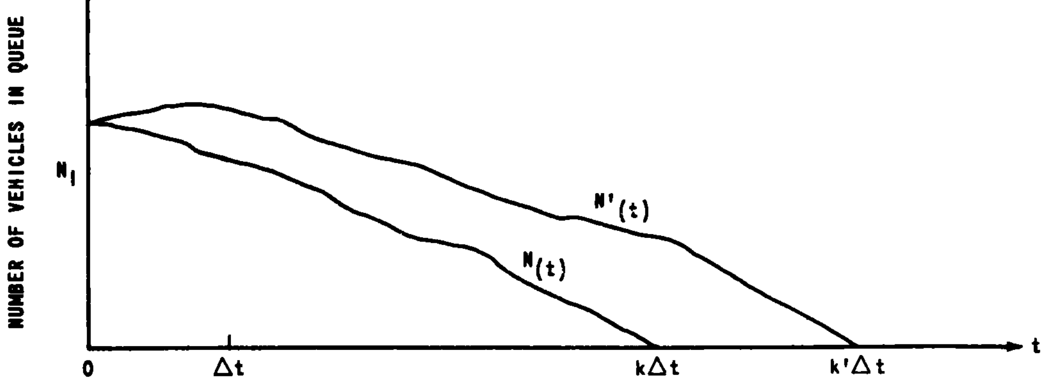


Figure D-2.

If $\theta \geq 0$, the signals should be left unchanged because less or equal delay would be incurred by an extension of the $\phi_{G|N,S}$, $\phi_{R|E,W}$ phase. However, if $\theta < 0$ a greater delay will occur if the signals are changed Δt later. This latter test is not sufficient to terminate the current phase immediately, because calculations of $\theta_{2\Delta t}$, $\theta_{3\Delta t}$, . . ., $\theta_{n\Delta t}$ may yield a θ-value algebraically greater than $\theta_{\Delta t}$. The mathematical and practical limitations on the number n that may be employed and produce meaningful results depends on the detectors (type and locations), current traffic characteristics, the estimators employed, and the upper limit $\phi_{R|E,W_{max}}$.

NOMENCLATURE AND DEFINITIONS

Static network = that portion of the network which may be considered fixed (i.e., those characteristics which may be stored in the computer's permanent memory, including number of lanes, turning prohibitions, street dimensions, etc.);

Dynamic network = that portion of the network which may be considered variable (i.e., those characteristics which must be measured "instantaneously" or predicted, including vehicular and pedestrian volumes, distribution of right and left turning vehicles, etc.);

$\phi_{G|E,W}$, $\phi_{G|N,S}$ = that phase which displays the green signal to the east-west and north-south legs of the intersection, respectively;

$\phi_{R|E,W}$, $\phi_{R|N,S}$ = that phase which displays the red signal to the east-west and north-south legs of the intersection, respectively;

$\phi_{G_{min}}$ = the minimum amount of time a green phase must receive before the phase may be changed by the computer's decision;

$\phi_{R_{max}}$ = the maximum amount of time a red phase may receive before it is changed by the computer, independent of current traffic demands;

ϕ_{tG}, ϕ_{tR} = elapsed time of the current green and red phases, respectively;

${}_{A}[\,]$ = with respect to intersection A;

Δt = unit of time considered for the proposed extension, and $1/\Delta t$ is the resultant (network) sampling rate;

$S_{T|E,W}$ = total delay saved to the east-west traffic;

S_D = a component of S_T, attributed to the expected departures;

S_q = a component of S_T, attributed to the number of vehicles serviced earlier in time;

d_E, d_W = departure rates, in vehicles per second, from the east and west legs, respectively, during the proposed extension, Δt;

$l_{E,W}$ = lost time attributed to the initial acceleration period to east (or west) vehicles;

ϕ_A = total amber phase, in seconds;

ϕ'_A = portion of the amber phase that may be considered an extension of the red phase;

$\sim$ = an estimate of the term to which it is applied;

a_{tE}, a_{tW} = arrival rates, in vehicles per second, on the east and west legs, respectively;

d_{L^+E} = departure rate, in vehicles per second, from the east legs during the initial portion of the next phase, $\phi_{G|E}$;

N_r = residual queue, in number of vehicles, on a particular leg at the change of phase;

$L = (\tilde{\phi}_R + \tilde{\phi}'_A + \tilde{l})/\Delta t$ calculated to the nearest whole number;

D_T = delay experienced to traffic on a particular leg resulting from the proposed extension of ϕ_R;

D_{dec} = a component of D_T, attributed to the vehicles decelerating to join end of queue;

D_{acc} = a component of D_T, attributed to those vehicles stopped by the queue which must accelerate to free speed;

D_q = a component of D_T, attributed to the delay suffered while in queue;

D_{norm} = a component of D_T, representing the vehicle-seconds required to traverse unimpeded the path to which the delay equation is applied;

d' = departure rate, not including the lost time (i.e., excluding acceleration (from stop) transient);

N_I = length of queue (number of vehicles) at the start of the new phase;

k, k' = integers representing the number of intervals, Δt, until the queues are effectively dissipated (for the two alternatives);

Δt_q = additional time (relative to the alternative) that the queue remains;

K_1, K_2 = constants of proportionality for D_{dec} and D_{acc} terms, respectively;

$N(t)$, $N'(t)$ = queue lengths (number of vehicles) for the cases where the phase is changed immediately or at an interval Δt later, respectively;

$a_{t'}$, $d_{t'}$, $\Delta t_{j'}$ = referring to $N'(t)$; and

$\theta_{\Delta t}$ = a quantity employed to test for minimum delay between the alternatives of an immediate change in signals or a postponement of Δt.

APPENDIX E

DERIVATION OF A MINIMUM-DELAY FUNCTION FOR A SUBNETWORK OF SIGNALIZED INTERSECTIONS

The minimum-delay equations developed for signalized intersection A in Appendix D do not account for the delay experienced downstream by those vehicles released from A. Although these isolated-intersection equations are applicable to all situations in which the state of the downstream network may be neglected, the vehicular arrival rates to A, a_i, may de *dependent* on a neighboring upstream signalized intersection (see Fig. E-1). In this case, the predictions, $\tilde{a}_i$, will reflect such influences.

The choice to neglect a certain portion of the surrounding network with respect to the decision to be made at intersection 2 (Fig. E-1) is a logical (practical) assumption and as was shown does not imply that such an intersection is literally isolated (i.e., does not influence or is not influenced by conditions at other intersections). It does, however, indicate that the intersection is being treated in such a manner, because the vehicles may have to travel relatively large distances before reaching another point (signalized intersection or measurement station), or pass through a number of unsignalized (non-instrumented) minor intersections before reaching a major intersection, or travel a relatively short distance where there may be a significant amount of friction (e.g., sinks and/or sources), or any combination of these conditions. It should be noted that in all the aforementioned instances it may be possible to say with a high degree of certainty that the vehicles will arrive at a particular point; but when in time (e.g., during which cycle or even which day) they will arrive has associated with it a very large uncertainty. In such instances it is apparently futile to link the intersection being evaluated with the neighboring ones by estimating the delay incurred at the signalized intersection downstream from the released vehicles. However, in urban areas, and in particular within the central business districts, such isolated intersections are in the minority. This is basically attributed to the geometry of the grid network (e.g., city street dimensions, and layout), and the vehicular and pedestrian volumes attracted to the area, which necessitates a high density of signalized intersections, and the more or less ordered flow of traffic which is prescribed by the enforcement of various traffic ordinances (e.g., parking, turning movements, speed limits, one-way designations).

A question which logically arises at this point is: How many signalized intersections downstream from the vehicles released from the intersection being evaluated, should be considered in the network delay equation? In the most general case of an intersection of two two-way streets (with no turning prohibitions) each vehicle has the choice of three paths (straight through, left, or right turns). As the number of intersections being considered is increased, the probability of the released vehicles reaching the farthest intersection downstream (in the original direction of travel) is decreased as a function of the distributions of the turning movements at each intersection. Further, one may examine the situation whereby turning movements are prohibited and the arrivals at a point downstream are certain, but are dispersed only by the variance on the mean velocity, which has associated with it the variance of the arrival times, accounting for the distance traveled. With the inclusion of one signal, or even one stop sign, controlling another intersection between the intersection being evaluated and the one downstream, the standard deviation of the predicted arrival times may be greater than the duration of the green phase (for the downstream signal), thus introducing a significant error in the predicted delay.

For practical reasons, such as those described, network equations will generally be limited to the immediate neighboring (minimum interference) signalized intersections, normally consisting of four pairs of intersections each containing the intersection being evaluated.

GENERAL PROBLEM STATEMENT

This appendix develops a set of general equations to measure the relative delay to vehicles currently being controlled at intersection q, that will include the predicted delay when these vehicles reach downstream intersections r, s, t, and u (Fig. E-2), for the control alternatives assumed to be available at q. Although the minimum-delay solution is sought in terms of the five intersections of two two-way streets, it is realized that many other network configurations are possible. However, the most common ones will simply be modifications of the foregoing subnetwork.

The problem is to decide on the "optimum" time to change the signal at q, assuming a symmetrical signal display, where "optimum" signifies the minimum aggregate delay accrued to those vehicles being controlled at intersection q.

The current conditions at q and the state of the rest of the network (e.g., particular phase, phase time elapsed, queue lengths, approach rates, departure rates, and various predictions based on past measurements) will be used to implement the minimum-delay criterion. The solution, for the most part, is based on the general development of the isolated-intersection model (Appendix D). Although the isolated case was concerned solely with the delay associated with intersection q, the equations are here being extended by considering the delays incurred in the subnetwork contained within the dashed lines of Figure E-2.

First let it be assumed that the signal at q is green to east- and westbound traffic (west and east legs, respectively); namely, phase ${}_q\phi_{G|E,W}$ (${}_q\phi_{R|N,S}$). Hence, it must be decided whether the current phase should be extended by $n\Delta t$ or terminated immediately (i.e., ${}_q\phi_{R|E,W}$, ${}_q\phi_{G|N,S}$

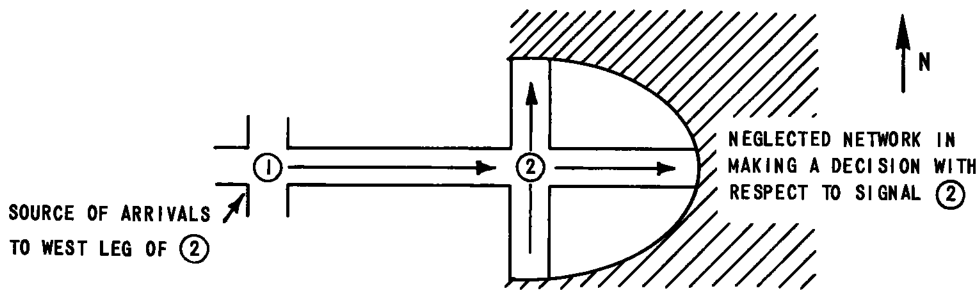


Figure E-1. Intersection of two one-way streets.

initiated). To evaluate the alternatives, the delay accrued by the vehicles up to the time that they depart from the subnetwork (at the perimeter of the dashed enclosure) must be considered. The implementation may take the form of evaluating either the difference in times of service at the perimeter (exit of subnetwork), or the sum of delays incurred at each intersection along the route. The latter approach is pursued here. Hence, the total relative delay or savings will be considered as the sum of contributions, from at most two signalized intersections for any particular vehicle. Therefore, the $\theta_{n\Delta t}$-term for the isolated intersection (Eq. D-19) is applicable to the total development and is the starting point.

SAVINGS WITHIN SUBNETWORK

The $S_{T|E,W}$ term (defined by Eq. D-7) is only a part of the total savings within the subnetwork, as this is the savings of delay at intersection q only. The additional relative savings within the entire subnetwork is expressed as the sum of the individual savings at each of the neighboring signalized intersections, as follows:

$$^{n\Delta t}{}_{q}S' = {}^{n\Delta t}({}_{q,r}S + {}_{q,s}S + {}_{q,t}S + {}_{q,u}S) \qquad \text{(E-1)}$$

In this notation, the first pre-subscript, q, refers to the situation whereby the savings equation compares the relative benefits of the alternative decisions to be made at intersection q; the second pre-subscript refers to the intersection within the subnetwork where the saving occurs. Further, the ${}_{q}S'$ term does not express the total* relative savings within the subnetwork q, r, s, t, u. This term is denoted by ${}^{n\Delta t}{}_{q}S$, where

$$^{n\Delta t}{}_{q}S = {}^{n\Delta t}({}_{q}S' + {}_{q}S_{T|E,W}) \qquad \text{(E-2)}$$

and ${}_{q}S_{T|E,W}$ is defined by Eq. D-7. It is desired, now, to express each of the contributions to ${}^{n\Delta t}{}_{q}S'$ in terms of the difference in delays (i.e., vehicle-seconds in addition to those necessary to travel through the subnetwork from q unimpeded) suffered at the neighboring intersections, which would occur if the vehicles are released as a result of the extension of the current green phase at q by $n\Delta t$ seconds,

* This value may be either positive or negative. If it is negative, more delay is predicted to be incurred by the vehicles from the east, west legs of q, at the neighboring intersections r, s, t, u, if released during the proposed extension, $n\Delta t$, than if held at q for the subsequent red phase, ${}_{q}\phi_{R|E,W}$, and then released. The term "savings" is always applied to the vehicles on the legs currently receiving the green signal.

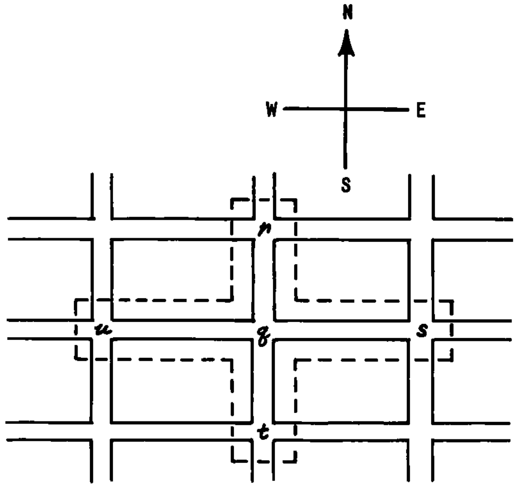


Figure E-2. Subnetwork (q, r, s, t, u) of signalized intersections.

or after being held for the duration of the next red phase, ${}_{q}\phi_{R|E,W}$. The expression for the q, r intersection pair is

$$^{n\Delta t}[{}_{q,r}S = {}_{q,r}D^{\tilde{\phi}_{R|E,W}} - {}_{q,r}D^{0}] \qquad \text{(E-3)}$$

where

${}_{q,r}D^{\tilde{\phi}_{R|E,W}}$ = the (estimated) delay incurred at r, to the vehicles released from intersection q after waiting for the next red phase, ${}_{q}\phi_{R|E,W}$, and choosing their path through r; and

${}_{q,r}D^{0}$ = the (estimated) delay incurred at r, to the same vehicles that are released during the current proposed extension and also desire service at r.

Similar equations may be written for the relative saving at s, t and u.

TURNING MOVEMENT CONSIDERATIONS

From Figure E-2 and the assumption of the initial state of q (i.e., $\phi_{G|E,W}$, $\phi_{R|N,S}$), it is readily observed that the candidates for service at r coming from q (to be considered in the term ${}_{q}S'$) must be contributions from the east and west legs, implying right- and left-turning movements, respectively. Hence, in order to predict the total relative savings, ${}_qS$, within the subnetwork, and the total relative delay, ${}_qD$, it is necessary to estimate the proportion (probability) of the vehicles that choose a particular direction from those which are available to them at q (i.e., the three alternatives, left turn, right turn, or straight through). The estimates of the proportions of turning movements* from each of the legs at q are defined as follows:

${}_q\tilde{x}_{iE}$ = estimate of the proportion of left turns from the east leg of intersection q during the ith interval;

${}_q\tilde{y}_{iE}$ = estimate of the proportion of right turns from the east leg of intersection q during the ith interval; and

${}_q\tilde{z}_{iE}$ = estimate of the proportion of straight-through's from the east leg of intersection q during the ith interval (i.e., ${}_q\tilde{z}_{iE} = 1 - {}_q(\tilde{x} + \tilde{y})_E$).

Similarly, one may define the estimates of these proportions at intersection q, from the west, north, and south legs.

The vehicles predicted to be "customers" for service at r (during the proposed extension of the ${}_q\phi_{G|E,W}$ phase) result from the sum of the left-turning movements from the west leg (eastbound) and the right-turning movements from the east leg (westbound). To arrive at this expression, the departure rate is multiplied by time, and also by the estimated proportion for the particular movement. Hence,

$$ {}^{n\Delta t}{}_{q,r}A^0(x,y) = \sum_{i=1}^{n} \{{}_q\tilde{x}_{iW}\, {}_q\tilde{d}_{iW}\, \Delta t_i\} + \sum_{i=1}^{n} \{{}_q\tilde{y}_{iE}\, {}_q\tilde{d}_{iE}\, \Delta t_i\} \quad \text{(E-4)} $$

where ${}^{n\Delta t}{}_{q,r}A^0(x,y)$ is the prediction of the number of arrivals at r resulting from the extension, $n\Delta t$, of the current signal phase at q (i.e., ${}_q\phi_{G|E,W}$).

In like manner the prediction of the number of arrivals to each of the neighboring intersections is obtained, as follows:

$$ {}^{n\Delta t}{}_{q,s}A^0(z) = \sum_{i=1}^{n} \{{}_q\tilde{z}_{iW}\, {}_q\tilde{d}_{iW}\, \Delta t_i\} \quad \text{(E-5)} $$

$$ {}^{n\Delta t}{}_{q,t}A^0(x,y) = \sum_{i=1}^{n} \{{}_q\tilde{x}_{iE}\, {}_q\tilde{d}_{iE}\, \Delta t_i\} + \sum_{i=1}^{n} \{{}_q\tilde{y}_{iW}\, {}_q\tilde{d}_{iW}\, \Delta t_i\} \quad \text{(E-6)} $$

$$ {}^{n\Delta t}{}_{q,u}A^0(z) = \sum_{i=1}^{n} \{{}_q\tilde{z}_{iE}\, {}_q\tilde{d}_{iE}\, \Delta t_i\} \quad \text{(E-7)} $$

where ${}^{n\Delta t}{}_{q,s}A^0(z)$, ${}^{n\Delta t}{}_{q,t}A^0(x,y)$, ${}^{n\Delta t}{}_{q,u}A^0(z)$ are the predictions of the number of arrivals at s, t, and u, respectively, resulting from the extension, $n\Delta t$, of the current signal phase at q (i.e., ${}_q\phi_{G|E,W}$). It should be noted that the predicted departure rate, $\tilde{d}$, is affected by the relative proportion of turning movement distributions, $\tilde{x}$, $\tilde{y}$, $\tilde{z}$, from both its own approach leg and the complementary direction.

* If, left-turning movements are prohibited by local traffic ordinances, ${}_q\tilde{x}_{iE} \equiv {}_q\tilde{x}_{iW} \equiv 0$ for all values of i.

SAVINGS EXPRESSED IN TERMS OF RELATIVE DELAYS

It is now necessary to examine the ${}_{q,r}D^{\tilde{\phi}_{R|E,W}}$ and ${}_{q,r}D^0$ terms of Eq. E-3. Referring to Eq. D-9, the required terms may be expressed as

$$ {}^{n\Delta t}{}_{q,r}D_S{}^{\tilde{\phi}_{R|E,W}} = {}^{n\Delta t}{}_{q,r}\{D_{dec} + D_q + D_{acc} - D_{norm}\}_S{}^{\phi_{R|E,W}} \quad \text{(E-8)} $$

and

$$ {}^{n\Delta t}{}_{q,r}D_S{}^0 = {}^{n\Delta t}{}_{q,r}\{D_{dec} + D_q + D_{acc} - D_{norm}\}_S{}^0 \quad \text{(E-9)} $$

in which the superscript and subscript notation follows the pattern ${}^1{}_{4,5}\{D_{8,3}\}_{6,7}{}^2$, wherein the various elements are as follows:

1 = decision interval, in seconds, currently being tested (i.e., extend current signal state by this value or terminate immediately), where the value may be Δt, $2\Delta t$, . . ., $n\Delta t$.

2 = time of start of next green interval as displayed to legs 6, 7, referenced from current instant, where the value may be 0 or ϕ_R (or ϕ_A and $n\Delta t$).

3 = leg at which delay, D, is accrued (i.e., N, S, E, or W).

4 = intersection at which decision is to be made on the state of the signal.

5 = intersection, within the subnetwork, at which the delay, D, is accrued.

6, 7 = legs at intersection (4) which contribute to ${}_4S$.

8 = component of delay (i.e., dec, acc, q, norm, or, if deleted, the sum of components).

Substituting Eqs. E-8 and E-9 in Eq. E-3 gives

$$ {}^{n\Delta t}{}_{q,r}S = {}^{n\Delta t}{}_{q,r}\{(D_{dec}{}^{\tilde{\phi}_{R|E,W}} - D_{dec}{}^0) + (D_q{}^{\tilde{\phi}_{R|E,W}} - D_q{}^0) + (D_{acc}{}^{\tilde{\phi}_{R|E,W}} - D_{acc}{}^0) - (D_{norm}{}^{\tilde{\phi}_{R|E,W}} - D_{norm}{}^0)\}_S \quad \text{(E-10)} $$

Inasmuch as the normal (unimpeded) time to travel from intersection q to intersection r remains unchanged with time, the D_{norm} terms are assumed to have negligible affect on the solution, because they will differ only with the variation in queue size (e.g., $D_{norm}{}^{\tilde{\phi}_{R|E,W}} \approx D_{norm}{}^0$). This assumption reduces the fourth term of Eq. E-10 to zero.

DELAY SUFFERED IN QUEUE

The second term on the right-hand side of Eq. E-10—namely, ${}_{q,r}(D_q{}^{\tilde{\phi}_{R|E,W}} - D_q{}^0)_S$—expresses the difference in delay suffered within queue on the south leg of intersection r. The vehicles considered are those predicted to be released from the east and west legs of intersection q during the current extension and the same vehicles whose departure is deferred by ${}_q\phi_{R|E,W}$. It is assumed that the number of vehicles predicted to demand service at r during the current proposed extension would have had the opportunity for service during the next green phase, ${}_q\phi_{G|E,W}$ (i.e., $n\Delta t \ll {}_q\tilde{\phi}_{G|E,W}$).

The $D_q{}^0$ term is then expressed in a form similar to the numerical integration equation (Eq. D-18). However, the period during which delay in the queue is computed in this instance, ${}_rK_{S,E,W}{}^0\Delta t$, begins when the first vehicle from q arrives at r and ends when the last vehicle arriving from q is released at r, or until the time of the queue's dissipation (whichever occurs first). Hence, the delay to the vehicles initially queued on the south leg of intersection r, at the time of the first arrival, should be subtracted from the total delay experienced in the same queue.

Therefore,

$$
{}^{n\Delta t}{}_{q,r}D_{q,S}{}^{0} = \sum_{j=1}^{K^0} {}_{q,r}\left\{\left[{}^{n\Delta t}\tilde{N}_{1S_j}{}^{0} + \sum_{i=1}^{j}\{(\tilde{a}'_i - \tilde{d}_i)\Delta t_i\} - (\tilde{a}'_j - \tilde{d}_j)\frac{\Delta t_j}{2}\right]\Delta t_j\right\}_S - \sum_{j=1}^{m^0} {}_{q,r}\left\{\left[\tilde{N}_{1S_j}{}^{0} - \sum_{i=1}^{j}(\tilde{d}_i\Delta t_i) + \tilde{d}_j\frac{\Delta t_j}{2}\right]\Delta t_j\right\}_S \quad \text{(E-11}a\text{)}
$$

in which a'_i, the arrivals at r, are considered to reflect only those contributions from intersection q during the immediate extension. Therefore, if $M^0\Delta t$ is defined as the interval of time over which departures from q during $n\Delta t$ arrive at r,

$$
\sum_{i=1}^{M^0} {}_{r}\tilde{a}_{iS}\Delta t_i = \sum_{i=1^-}^{n^-} {}_{q}\{(\tilde{x}_{iW}\,\tilde{d}_{iW}\,\Delta t_i) + (\tilde{y}_{iE}\,\tilde{d}_{iE}\,\Delta t_i)\} \quad \text{(E-11}b\text{)}
$$

where the summations yield the same number of vehicles, although the summations occur at different times as denoted by the minus superscripts on the limits (i.e., the departures from intersection q occur earlier than their subsequent arrivals at r). Further, in general, $M^0 \neq n$, which allows for the variation in velocities between first and last vehicles considered and also for the fluctuating queue length on the south leg of intersection r.

Therefore the following conditional expressions can be written:

If

$$
1 \leq i \leq M^0 \quad \text{(E-12)}
$$

then

$$
{}_{r}a'_{iS} \equiv {}_{r}a_{iS} \quad \text{(E-13)}
$$

If

$$
i > M^0 \quad \text{(E-14)}
$$

then

$$
{}_{r}a'_{iS} \equiv 0 \quad \text{(E-15)}
$$

Substituting average values for the individual estimates obtained per time interval in Eq. E11b gives

$$
{}_{r}\bar{a}_{iS}M^0\Delta t = (\overline{x_{iW}d_{iW} + y_{iE}d_{iE}})n\Delta t \quad \text{(E-16)}
$$

where the bar notation implies the average value of the particular quantity, taken over the intervals $M^0\Delta t$ and $n\Delta t$, respectively, and it is assumed that the interval Δt_i is invariant with respect to i. Solving for M^0 gives

$$
M^0 = \frac{(\overline{x_{iW}d_{iW} + y_{iE}d_{iE}})n}{{}_{r}\bar{a}_{iS}} \quad \text{(E-17)}
$$

The upper limit, m^0, for the second part of Eq. E-11a, is defined as follows:

$m^0\Delta t$ is the predicted interval of time necessary to serve vehicles originally queued on the south leg of r, referenced from the instant that the first vehicle arriving from intersection q (as a result of the current extension) joins the existing queue at r, until this existing queue is served.

Since ${}_{r}\tilde{N}_S{}^0$ is defined as the number of vehicles in queue on the south leg of intersection r at the time of the first arrival from intersection q,

$$
{}_{r}N_{1S}{}^{0} - \sum_{i=1}^{m^0} {}_{r}\tilde{d}_{iS}\Delta t_i = 0 \quad \text{(E-18)}
$$

in which the departures from the south leg of intersection r, ${}_{r}d_{iS}\Delta t_i$, are considered to be summed from the same lower limit $(i = 1)$ as the summation of the arrivals, ${}_{r}a_{iS}\Delta t_i$, in Eq. E-13 were. Substituting average values for the estimates and rearranging terms gives

$$
{}_{r}N_{1S}{}^{0} = {}_{r}\bar{d}_{iS}m^0\Delta t \quad \text{(E-19)}
$$

and

$$
m^0 = \frac{{}_{r}N_{1S}{}^{0}}{{}_{r}\bar{d}_{iS}\Delta t} \quad \text{(E-20)}
$$

The ${}_{r}K_{S|E,W}{}^{0}$ limit of Eq. E-11a is now examined in detail. First, ${}_{r}K_{S|E,W}{}^{0}\Delta t$ is defined as the interval of time, referenced from the time of the first arrival to r's south leg, until the queue on this leg is dissipated* or until the last vehicle being considered from the east-west extension at q is serviced at r, whichever event occurs first. Second, ${}_{r}K_{S|E,W}{}^{0'}\Delta t$ is defined as the interval of time, referenced from the time of the first arrival to r's south leg, until the queue on this leg is dissipated. Then

$$
{}^{n\Delta t}{}_{r}N_{1S}{}^{0} + \sum_{i=1}^{{}_{r}K_{S|E,W}{}^{0'}} {}_{r}(\tilde{a}_{iS} - \tilde{d}_{iS})\Delta t_i = 0 \quad \text{(E-21)}
$$

and substitution of average values for the estimates gives

$$
{}^{n\Delta t}{}_{r}N_{1S}{}^{0} + {}_{r}(\bar{a}_i - \bar{d}_i)_S\;{}_{r}K_{S|E,W}{}^{0'}\Delta t = 0 \quad \text{(E-22)}
$$

from which

$$
{}_{r}K_{S|E,W}{}^{0'} = \frac{-{}_{r}N_{1S}{}^{0}}{{}_{r}(\bar{a} - \bar{d})_S\Delta t} \quad \text{(E-23)}
$$

If ${}_{r}K_{S|E,W}{}^{0''}\Delta t$ is defined as the interval of time, referenced from the time of the first arrival to r's south leg, in which the last vehicle being considered from the east-west extension at q is serviced at r,

$$
{}_{r}N_{1S}{}^{0} + \sum_{i=1}^{m^0} {}_{r}\tilde{a}_{iS}\Delta t_i - \sum_{i=1}^{{}_{r}K_S{}^{0''}} {}_{r}\tilde{d}_{iS}\Delta t_i = 0 \quad \text{(E-24)}
$$

and substitution of average values for the estimates gives

$$
{}^{n\Delta t}{}_{r}N_{1S}{}^{0} + {}_{r}\bar{a}_S m^0\Delta t - {}_{r}\bar{d}_S\;{}_{r}K_{S|E,W}{}^{0''}\Delta t = 0 \quad \text{(E-25)}
$$

from which

$$
{}_{r}K_{S|E,W}{}^{0''} = \frac{{}^{n\Delta t}{}_{r}N_{1S}{}^{0} + {}_{r}\bar{a}_S m^0\Delta t}{{}_{r}\bar{d}_S\Delta t} \quad \text{(E-26)}
$$

Now, recalling the definitions following Eq. E-20, the following conditional equalities may be written:

* It is assumed that the level of service at r with respect to the demand on the south leg will be such that once the queue is dissipated no further build-up will occur during the time in which the remainder of the vehicles under consideration arrive.

If

$$ {}_{\mathrm{r}}K_{\mathrm{S|E,W}}{}^{0'} \leqq {}_{\mathrm{r}}K_{\mathrm{S|E,W}}{}^{0''} \tag{E-27} $$

then

$$ {}_{\mathrm{r}}K_{\mathrm{S|W,E}}{}^{0} \equiv {}_{\mathrm{r}}K_{\mathrm{S|E,W}}{}^{0'} \tag{E-28} $$

However, if

$$ {}_{\mathrm{r}}K_{\mathrm{S|E,W}}{}^{0'} > {}_{\mathrm{r}}K_{\mathrm{S|E,W}}{}^{0''} \tag{E-29} $$

then

$$ {}_{\mathrm{r}}K_{\mathrm{S|E,W}}{}^{0} \equiv {}_{\mathrm{r}}K_{\mathrm{S|E,W}}{}^{0''} \tag{E-30} $$

Now that the delay suffered in queue on the south leg of r has been estimated for the alternative of extending the current green phase, a prediction can be made of the delay that would be incurred in queue to these same vehicles at r, after being held for the next amber and red phase $(\phi'_{\mathrm{A|E,W}} + \phi_{\mathrm{R|E,W}})$; namely, ${}_{\mathrm{r}}D_q{}^{\tilde{\phi}_{\mathrm{R|E,W}}}$. The expression, similar in form to Eq. E-11*a*, is

$$ {}^{n\Delta t}{}_{\mathrm{q,r}}D_{q\mathrm{S|E,W}}{}^{\phi_{\mathrm{R}}} = \sum_{j=1+a}^{({}_{\mathrm{r}}K_{q\mathrm{S|E,W}}{}^{\phi_{\mathrm{R}}+a})} {}_{\mathrm{r}}\left\{\left[N_{1\mathrm{S}_j}{}^{\phi_{\mathrm{R|E,W}}} + \sum_{i=1+a}^{j} \{(\tilde{a}'_i - \tilde{d}_i)\Delta t_i\} - (\tilde{a}'_j - \tilde{d}_j)\frac{\Delta t_j}{2}\right]\Delta t_j\right\}_{\mathrm{S}} - \sum_{j=1+a}^{m^{\phi_{\mathrm{R}}}+a} {}_{\mathrm{r}}\left\{\left[N_{1\mathrm{S}_j}{}^{\phi_{\mathrm{R|E,W}}} - \sum_{i=1}^{j} (\tilde{d}_i\Delta t_i) + \tilde{d}_j\frac{\Delta t_j}{2}\right]\Delta t_j\right\}_{\mathrm{S}} \tag{E-31} $$

in which the lower limit on the summations indicates that the delays to the vehicles are considered later in time than Eq. E-11*a*, and $a\Delta t$ is defined as the interval of time which is estimated to elapse between the first arrival on the south leg of r due to the extension of the current phase at q and the first arrival from the same legs at q due to the termination of the current phase (i.e., deferred in time).

To clarify this definition and develop an expression relating a to defined physical quantities, the following discussion refers to Figures E-3 and E-4.

The shaded vehicle is the same one in both Figure E-3 and Figure E-4. However, it is released $(\tilde{\phi}_{\mathrm{R'E,W}} + \phi'_{\mathrm{A}})$ seconds later in Figure E-4 than in Figure E-3. Assuming an average velocity from the time of departing intersection 1 to falling in behind the last vehicle in the queue, the travel time to fall in place in queue for the two alternatives may be expressed as b/V_{avg} and b'/V'_{avg}, respectively, where V_{avg} and V'_{avg} represent the average velocity within b and b', respectively, and b and b' are the respective distances from intersection 1 to the end of the queue at the time of the first vehicle's falling in behind the last vehicle. Hence, the delay expressed in queue, $D^{\phi_{\mathrm{R}}}$, will apply $\left\{\phi_{\mathrm{R}} + \phi_{\mathrm{A}} + \frac{b'}{V'_{\mathrm{avg}}} - \frac{b}{V_{\mathrm{avg}}}\right\}$ seconds later than D^0, and

$$ a \equiv \frac{\left\{\phi_{\mathrm{R}} + \phi_{\mathrm{A}} + \left(\frac{b'}{V'_{\mathrm{avg}}} - \frac{b}{V_{\mathrm{avg}}}\right)\right\}}{\Delta t} \tag{E-32} $$

By expressing relationships analogous to Eqs. E-16, E-18 and E-19, E-21 and E-22, and E-24 and E-25, the following parameters may be arrived at, respectively:

$$ M^{\phi_{\mathrm{R}}} = \frac{(\overline{x_{i\mathrm{W}}d_{i\mathrm{W}} + y_{i\mathrm{E}}d_{i\mathrm{E}}})n}{{}_{\mathrm{r}}\bar{a}_{i\mathrm{S}}} \tag{E-33} $$

$$ m^{\phi_{\mathrm{R}}} \doteq \frac{{}_{\mathrm{r}}N_{1\mathrm{S}}{}^{\phi_{\mathrm{R}}}}{{}_{\mathrm{r}}\bar{d}_{i\mathrm{S}}\Delta t} \tag{E-34} $$

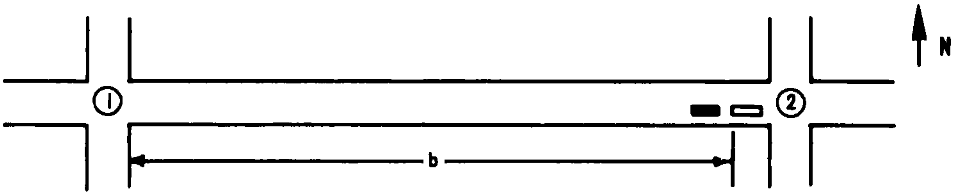


Figure E-3. Prediction of length of queue at point 2 (west leg) b/V$_{\mathrm{avg}}$ *seconds after decision to extend* $\phi_{\mathrm{G\,E\,W}}$.

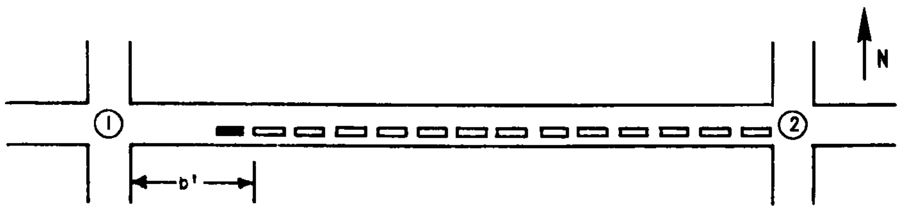


Figure E-4. Prediction of length of queue at point 2 (west leg) b'/V'$_{\mathrm{avg}}$ *seconds after start of deferred* $\phi_{\mathrm{G|E,W}}$.

$$ {}_{\mathrm{r}}K_{\mathrm{S|E,W}}{}^{\phi'_{\mathrm{R}}} = \frac{-\,{}_{\mathrm{r}}N_{1\mathrm{S}}{}^{\phi_{\mathrm{R}}}}{{}_{\mathrm{r}}(\bar{a}_1 - \bar{d}_i)_{\mathrm{S}}\Delta t} \quad \text{(E-35)} $$

$$ {}_{\mathrm{r}}K_{\mathrm{S|E,W}}{}^{\phi''_{\mathrm{R}}} = \frac{{}_{1}N_{1\mathrm{S}}{}^{\phi_{\mathrm{R}}} + {}_{i}\bar{a}_{i\mathrm{S}}m^{\phi_{\mathrm{R}}}\Delta t}{{}_{\mathrm{r}}\bar{d}_{\mathrm{S}}\Delta t} \quad \text{(E-36)} $$

where the time averages are assumed to be of the estimates per interval over the appropriate (time-deferred) interval.

DECELERATION AND ACCELERATION TERMS

The next step is to return to the first and third terms of Eq. E-10 and investigate the delay incurred by the vehicles being considered during their deceleration and acceleration states.

Two cases are examined. First, if there is no queue predicted to be formed on the south leg of intersection r, at either time of the two alternatives, the relative deceleration and acceleration delay terms will be zero. This implies that the relative level of service between q and r is such that no queue exists during the entire time over which arrivals are considered at r.

Second, if a queue persists during both alternatives, there will again be no relative advantage by virtue of the D_{acc} and D_{dec} considerations on the south leg of r.

Hence, the delay consequences of both the cases described may be expressed as

$$ (D_{\mathrm{dec}}{}^{\phi_{\mathrm{R}}} - D_{\mathrm{dec}}{}^{0}) = (D_{\mathrm{acc}}{}^{\phi_{\mathrm{R}}} - D_{\mathrm{acc}}{}^{0}) = 0 \quad \text{(E-37)} $$

which is based on the assumption of uniform deceleration and acceleration characteristics (refer to Eqs. D-14 and D-15).

The next step is to estimate the number of vehicles of the total considered, ${}_{\mathrm{q,r}}A^0(x,y)$, that must decelerate (and accelerate) due to the formation of a queue on the south leg of r (or in the case of the first vehicle, the ${}_{1}\phi_{\mathrm{R|N,S}}$ phase) during the two alternatives.

Referring to Eq. E-22, it will be recalled that ${}_{\mathrm{r}}K^{0'}\Delta t$ represents the interval of time measured from the first arrival until the dissipation of the queue. Since arrivals at r are assumed to take place over ${}_{\mathrm{r}}M^0\Delta t$ (see Eq. E-16), referenced from the same event, the following conditional relationships may be written:

If

$$ K^{0'} \geq M^0 \quad \text{(E-38)} $$

the queue persists during the entire time of the ${}_{\mathrm{q,r}}A^0(x,y)$ arrivals. Hence,

$$ {}^{n\Delta t}{}_{\mathrm{q,r}}D_{\mathrm{dec\ S}}{}^{0} = {}_{\mathrm{r}}\left(K_1 \sum_{i=1}^{M^0} \tilde{a}_i \Delta t_i\right)_{\mathrm{S}} \quad \text{(E-39)} $$

and

$$ {}^{n\Delta t}{}_{\mathrm{q,r}}D_{\mathrm{acc}}{}^{0} = {}_{\mathrm{r}}\left(K_2 \sum_{i=1}^{M^0} \tilde{a}_i \Delta t_i\right)_{\mathrm{S}} \quad \text{(E-40)} $$

However, if

$$ K^{0'} < M^0 \quad \text{(E-41)} $$

only a portion of the ${}_{\mathrm{q,r}}A^0(x,y)$ vehicles considered in the general delay equation incur the delay, as follows:

$$ {}^{n\Delta t}{}_{\mathrm{q,r}}D_{\mathrm{dec}}{}^{0} = {}_{\mathrm{r}}\left(K_1 \sum_{i=1}^{K^{0'}} \tilde{a}_i \Delta t_i\right)_{\mathrm{S}} \quad \text{(E-42)} $$

$$ {}^{n\Delta t}{}_{\mathrm{q,r}}D_{\mathrm{acc\ S}}{}^{0} = {}_{\mathrm{r}}\left(K_2 \sum_{i=1}^{K^{0'}} \tilde{a}_i \Delta t_i\right)_{\mathrm{S}} \quad \text{(E-43)} $$

In the case of Eqs. E-42 and E-43 the queue is assumed to be dissipated over the interval $(M^0 - K^{0'})\Delta t$. Therefore, the arrivals during this time $\left\{\sum_{K^{0'}}^{M^0} {}_{\mathrm{r}}\tilde{a}_{i\mathrm{S}}\Delta t_i\right\}$ are assumed to accrue zero delay in terms of $D_{\mathrm{dec}}{}^{0}$, $D_{\mathrm{acc}}{}^{0}$ and $D_{q}{}^{0}$.

Equations analogous to E-39 through E-43 may be written for the alternative of terminating the current phase at intersection q. Hence, the delays would be predicted for the same vehicles, under the conditions that are anticipated after the vehicles are required to wait for the subsequent red phase, ${}_{\mathrm{q}}\phi_{\mathrm{R|E,W}}$. These quantities are denoted by the previously deferred $D_{\mathrm{acc}}{}^{\tilde{\phi}_{\mathrm{R|E,W}}}$, $D_{\mathrm{dec}}{}^{\tilde{\phi}_{\mathrm{R|E,W}}}$ terms.

DELAY PREDICTION

Up to this point consideration has been given only to the "savings" of delay, ${}_{\mathrm{q}}S$, within the subnetwork for those same vehicles predicted to receive service at intersection q (east-west legs) during the alternatives available at q (0, $\tilde{\phi}_{\mathrm{R|E,W}}$), by comparing the delay experienced downstream at the neighboring intersections.

An attempt is next made to define the expression for the relative delay*, ${}_{\mathrm{q}}D$, caused within the subnetwork, in terms of the previously defined delay at q, $D_{T\mathrm{S}}$ and $D_{T\mathrm{N}}$ (isolated case, see Eq. D-9), and the relative delay accrued to the vehicles at the neighboring intersections which are predicted to be released from the (opposite) north and south legs of q.

Here, the predicted consequences of the two alternative actions at intersection q must again be compared, as follows:

1. Terminate current phase, initiate ${}_{\mathrm{q}}\phi_{\mathrm{G|N,S}}$ now, or
2. Defer action 1 by the extension $n\Delta t$ of ${}_{\mathrm{q}}\phi_{\mathrm{G|E,W}}$.

First, the delay caused within the subnetwork is expressed, as follows:

$$ {}^{n\Delta t}{}_{\mathrm{q}}D = {}^{n\Delta t}{}_{\mathrm{q}}D' + {}_{\mathrm{q}}D_{T|\mathrm{N,S}} \quad \text{(E-44)} $$

where, from Appendix D,

$$ {}^{n\Delta t}{}_{\mathrm{q}}D_{T|\mathrm{N,S}} = {}^{n\Delta t}({}_{\mathrm{q}}D_{T\mathrm{S}} + {}_{\mathrm{q}}D_{T\mathrm{N}}) \quad \text{(E-45)} $$

and

$$ {}^{n\Delta t}{}_{\mathrm{q}}D' = {}^{n\Delta t}({}_{\mathrm{q,r}}D + {}_{\mathrm{q,s}}D + {}_{\mathrm{q,t}}D + {}_{\mathrm{q,u}}D) \quad \text{(E-46)} $$

As previously, the first pre-subscript indicates the intersection to which the decision is applied and the second indicates the particular intersection at which the relative delay is considered.

The next choice is to investigate a particular intersection—namely, r—within the subnetwork q, r, s, t, u.

* Although this term is denoted as "relative delay," it should be noted that its sign may be positive or negative. If it is negative, vehicle-seconds are predicted to be saved within the subnetwork by the north-south departures if deferred an additional $n\Delta t$ seconds.

DELAY IN QUEUE

By referring to Eq. E-10, the relative delay in queue at r is written as

$$ {}_{q,r}D_q = {}_{q,r}(D_q{}^{n\Delta t} - D_q{}^{\phi_{A|E,W}})_{S_{N,S}} \qquad \text{(E-47)} $$

where

${}_{q,r}D_{qS_{N,S}}{}^{n\Delta t}$ = delay suffered within the queue on the south leg of r to those vehicles released from the north-south legs of intersection q if the current phase, ${}_q\phi_{G|E,W}$, is extended by $n\Delta t$; and

${}_{q,r}D_{qS_{N,S}}{}^{\phi_{A|E,W}}$ = delay suffered within the queue on the south leg of r to the vehicles released from the north-south legs of intersection q if the current phase, ${}_q\phi_{G|E,W}$, is terminated immediately (i.e., $\phi_{G|N,S}$ initiated $\phi_{A|E,W}$ later).

The D^{ϕ_A} term, which is similar in form to Eq. E-11, is now examined, as follows:

$$ {}^{n\Delta t}{}_{q,r}D_{qS_{N,S}}{}^{\phi_A} = \sum_{j=1+\beta}^{K\phi_A+\beta} \left\{ \left[\tilde N_j{}^{\phi_A} + \sum_{i=1+\beta}^{j} \{(\tilde a'_i - \tilde d_i)\Delta t_i\} - (\tilde a'_j - \tilde d_j)\frac{\Delta t_j}{2} \right] \Delta t_j \right\}_S - \sum_{j=1+\beta}^{m\phi_A+\beta} {}_r\left\{ \left[\tilde N_j{}^{\phi_A} - \sum_{i=1+\beta}^{j} (\tilde d_i \Delta t_i) + \tilde d_j \frac{\Delta t_j}{2} \right] \Delta t_j \right\}_S \qquad \text{(E-48)} $$

in which the summation limits reflect that the delay to the vehicles are considered later in time than Eq. E-11, as defined by:

$\beta\Delta t$ = interval of time which is estimated to elapse between the first arrival on the south leg of r, due to the extension of the current phase at q, and the first arrival from the opposite pair of legs at q, due to initiation of the new phase.

Hence, by making reference to the discussion immediately preceding Eq. E-32, the following may be written:

$$ \beta\Delta t = \phi_{A|E,W} \pm |\text{ difference in travel time }| \qquad \text{(E-49)} $$

The following restrictions and definitions apply to Eq. E-48.

The arrivals at r, a'_i, considered in Eq. E-48 should reflect only those contributions from the north-south legs of the newly initiated ${}_q\phi_{G|N,S}$ phase.

Therefore, if $M^{\phi_A}\Delta t_i$ is defined as the interval of time over which departures occurring during the entire $\phi_{G|N,S}$ phase arrive at r,

$$ \sum_{i=1+\beta}^{M\phi_A+\beta} {}_r\tilde a_{iS}\Delta t_i = \sum_{i=1+\beta^-}^{B+\beta^-} {}_q\{\tilde z_i \tilde d_i \Delta t_i\}_S \qquad \text{(E-50)} $$

where $B\Delta t$ is the estimate of the effective green time of the new phase (as displayed to the north-south legs). Therefore,

$$ B = \frac{\tilde\phi_{G|N,S} + (\tilde\phi_{A|N,S} - \phi'_{A|N,S})}{\Delta t} \qquad \text{(E-51)} $$

and the following conditional expressions may be written: If

$$ 1 + \beta \leq i \leq M^{\phi_A} + \beta \qquad \text{(E-52)} $$

then

$$ {}_r a'_{iS} = {}_r a_{iS} \qquad \text{(E-53)} $$

However, if

$$ i > M^{\phi_A} + \beta \qquad \text{(E-54)} $$

then

$$ {}_r a'_{iS} = 0 \qquad \text{(E-55)} $$

Substitution of average values for the estimates in Eq. E-50 gives

$$ M^{\phi_A} = \frac{{}_q\bar z_{iS}\, {}_q\bar d_{iS}\, B}{{}_r\bar a_{iS}} \qquad \text{(E-56)} $$

Similarly, the upper limit m^{ϕ_A}, which appears in the second part of Eq. E-48, may be defined as

$m^{\phi_A}\Delta t$ = predicted interval of time necessary to serve vehicles originally queued on the south leg of r, referenced from the instant that the first vehicle arriving from intersection q (as a result of the immediate phase change), joins the existing queue at r, until this queue is served.

Since ${}_r\tilde N_{1S}{}^{\phi_A}$ is defined as the number of vehicles in queue on the south leg of intersection r at the time of the first arrival from intersection q (resulting from the aforementioned decision at q), the following relationship may be written:

$$ {}_r\tilde N_{1S}{}^{\phi_A} - \sum_{i=1+\beta}^{m\phi_A+\beta} {}_r\tilde d_{iS}\Delta t_i = 0 \qquad \text{(E-57)} $$

Substitution of average values for the estimates and solution for m^{ϕ_A} gives

$$ m^{\phi_A} = \frac{{}_r\tilde N_{1S}{}^{\phi_A}}{{}_r\bar d_{1S}\Delta t} \qquad \text{(E-58)} $$

Similarly, one may solve for ${}_rK_{S|N,S}{}^{\phi_A\prime}$ and ${}_rK_{S|N,S}{}^{\phi_A\prime\prime}$ in order to choose $K_{S|N,S}{}^{\phi_A}$. Because these estimates are defined (following Eq. E-20) for the case to the 0*th* power east-west departures from q, the results are simply stated in terms of average values, as follows:

$$ {}_rK_{S|N,S}{}^{\phi_A\prime} = \frac{-\,{}_rN_{1S}{}^0}{{}_r(\bar a_i - \bar d_i)_S\Delta t} \qquad \text{(E-59)} $$

and

$$ {}_rK_{S|N,S}{}^{\phi_A\prime\prime} = \frac{{}_rN_{1S}{}^{\phi_A} + {}_r\bar a_{iS} m^{\phi_A}\Delta t}{{}_r\bar d_{iS}\Delta t} \qquad \text{(E-60)} $$

where ${}_rK_{S|N,S}{}^{\phi_A}$ is chosen in accordance with tests analogous to those performed in Eq. E-27 through E-30.

The delay suffered in queue, ${}_{q,r}D_{qN,S}{}^{n\Delta t}$, at r to these same vehicles if the start of the ${}_q\phi_{G|N,S}$ phase is deferred by an additional $n\Delta t$ seconds, may now be expressed as

$$_{q,r}D_{qS|N,S}{}^{n\Delta t} = \sum_{j=1+\gamma}^{{}_rK_S{}^{n\Delta t}+\gamma} \left\{\left[N_\gamma{}^{n\Delta t} + \sum_{i=1+\gamma}^{j}\{(\tilde{a}'_i - \tilde{d}_i)\Delta t_i\} - (\tilde{a}'_j - \tilde{d}_j)\frac{\Delta t_j}{2}\right]\Delta t_j\right\}_S - \sum_{j=1+\gamma}^{m^{n\Delta t}+\gamma} {}_r\left\{\left[N_j{}^{n\Delta t} - \sum_{i=1+\gamma}^{j}(\tilde{d}_i\Delta t_i) + \frac{\tilde{d}_j\Delta t_j}{2}\right]\Delta t_j\right\}_S \quad \text{(E-61)}$$

Inasmuch as the restrictions and definitions of terms used in Eq. E-61 are analogous to the preceding, the details are excluded and only the results are given, as follows:

$$\gamma = \beta + n \quad \text{(E-62)}$$

$$M^{n\Delta t} = \frac{\bar{z}_{iS}\bar{d}_{iS}B}{{}_r\bar{a}_{iS}} \quad \text{(E-63)}$$

(Inasmuch as the same number of arrivals, ${}_rA^{\phi_A}(z)$, is considered in each instance (ϕ_A or $n\Delta t$), it was assumed in Eq. E-63 that the duration of the deferred ${}_q\phi_{G|N,S}$ phase is at least as long as the one considered in Eq. E-51 with equivalent departure characteristics; if this is not the case, Eq. E-63 would be modified to reflect the same number of departures over a shorter or longer period.)

$$m^{n\Delta t} = \frac{{}_rN_{1S}{}^{n\Delta t}}{{}_r\bar{d}_{iS}\Delta t} \quad \text{(E-64)}$$

$${}_rK_{S|N,S}{}^{n\Delta t'} = \frac{-{}_rN_{1S}{}^{n\Delta t}}{{}_r(\bar{a} - \bar{d})_S\Delta t} \quad \text{(E-65)}$$

$${}_rK_{S|N,S}{}^{n\Delta t''} = \frac{{}_rN_{1S}{}^{n\Delta t} + {}_r\bar{a}_{iS}m^{n\Delta t}\Delta t}{{}_r\bar{d}_{iS}\Delta t} \quad \text{(E-66)}$$

DECELERATION AND ACCELERATION TERMS

One may now write the delay to those vehicles of the total considered, ${}_{q,r}A^{\phi_A}(z)$, that must decelerate (accelerate) due to conditions at r, in an analogous fashion to that expressed in Eqs. E-38 to E-43. Therefore, if

$$K^{n\Delta t} \geq M^{n\Delta t} \quad \text{(E-67)}$$

then

$$_{q,r}D_{dec}{}^{n\Delta t} = {}_r\left(K_1 \sum_{i=1+\gamma}^{m^{n\Delta t}+\gamma} \tilde{a}_i\Delta t_i\right)_S \quad \text{(E-68)}$$

and

$$_{q,r}D_{acc}{}^{n\Delta t} = {}_r\left(K_2 \sum_{i=1+\gamma}^{m^{n\Delta t}+\gamma} \tilde{a}_i\Delta t_i\right)_S \quad \text{(E-69)}$$

However, if

$$K^{n\Delta t} < M^{n\Delta t} \quad \text{(E-70)}$$

then

$$_{q,r}D_{dec}{}^{n\Delta t} = {}_r\left(K_1 \sum_{i=1+\gamma}^{K^{n\Delta t}+\gamma} \tilde{a}_i\Delta t_i\right)_S \quad \text{(E-71)}$$

and

$$_{q,r}D_{acc}{}^{n\Delta t} = {}_r\left(K_2 \sum_{i=1+\gamma}^{K^{n\Delta t}+\gamma} \tilde{a}_i\Delta t_i\right)_S \quad \text{(E-72)}$$

Similarly, for the alternative of releasing the vehicles from the legs currently receiving the red phase as soon as possible (after $\phi_{A|E,W}$),
if

$$K^{\phi'_A} \geq M^{\phi_A} \quad \text{(E-73)}$$

then

$$_{q,r}D_{dec}{}^{\phi_A} = {}_r\left(K_1 \sum_{i=1+\beta}^{M^{\phi_A}+\beta} \tilde{a}_i\Delta t_i\right)_S \quad \text{(E-74)}$$

and

$$_{q,r}D_{acc}{}^{\phi_A} = {}_r\left(K_2 \sum_{i=1+\beta}^{M^{\phi_A}+\beta} \tilde{a}_i\Delta t_i\right)_S \quad \text{(E-75)}$$

However, if

$$K^{\phi'_A} < M^{\phi_A} \quad \text{(E-76)}$$

then

$$_{q,r}D_{dec}{}^{\phi_A} = {}_r\left(K_1 \sum_{i=1+\beta}^{K^{\phi'_A}+\beta} \tilde{a}_i\Delta t_i\right)_S \quad \text{(E-77)}$$

and

$$_{q,r}D_{acc}{}^{\phi_A} = {}_r\left(K_2 \sum_{i=1+\beta}^{K^{\phi'_A}+\beta} \tilde{a}_i\Delta t_i\right)_S \quad \text{(E-78)}$$

GENERALIZING EQUATIONS

Although the particular equations were only developed for intersection r of the subnetwork, the extensions of these equations to s, t and u are obvious and were assumed without further explicit presentation. Further, in order to more readily develop the concepts, a particular initial condition was chosen (i.e., state of signal display at q, ${}_q\phi_{G|E,W}$, ${}_q\phi_{R|N,S}$). Because a number of possible signal states exist at an intersection, the following discussion and Table E-1 have been included for purposes of clarification and review. In Table E-1, Col. 1 specifies the particular pair of legs at q for each possible signal state; Col. 2 gives the signal displayed to each of the two pairs of legs at intersection q; Cols. 3 and 4 show that the "savings" equation (as expressed for the neighboring intersection r) previously developed (refer to Eq. E-3), applies to the vehicles on the north and south legs for signal state I and to the vehicles on the east and west legs for signal display II; Col. 5 indicates the total number and source of the vehicles to be considered in the calculation of Cols. 3 and 4 (that is, in case I the vehicles predicted to incur relative savings at r are departures from the south leg of q over period $n\Delta t$ that choose a northbound (straight through) direction, whereas in case II left- and right-turning movements are contributions as indicated); Cols. 6 and 7 show that the delay equations (as expressed for the neighboring intersection r) previously developed (refer to Eq. E-47 for delay in queue) apply to the vehicles on the east and west legs for case I and the north and south legs for case II; and Col. 8 gives the total number of vehicles and their source which are considered in the calculations of Cols. 6 and 7.

TABLE E-1

APPLICATION OF DERIVED EQUATIONS TO VARIOUS SIGNAL STATES AT q

SIGNAL STATE	LEG (PAIR) (*1*)	SIGNAL STATE AT q (*2*)	D^{ϕ_R} (*3*)	D^0 (*4*)	$^{n\Delta t}{}_rA^0(x, y, z)$ (*5*)	$_rD^{n\Delta t}$ (*6*)	$_rD^{\phi_A}$ (*7*)	$_rA^{\phi_A}(x, y, z)$ (*8*)
I	E, W	$_q\phi_{R\|E,W}$			$^{n\Delta t}{}_rA^0(z_S) = \sum_{i=1}^{n}{}_q\{\tilde{z}_{iS}\tilde{d}_{iS}\Delta t_i\}$	$_rD^{n\Delta t}$	$D^{\phi_A\|N,S}$	$_rA^{\phi_A}(x_W, y_E) = \sum_{i=1}^{B}{}_q\{(\tilde{x}_{iW}\tilde{d}_{iW}\Delta t_i) + (\tilde{y}_{iE}\tilde{d}_{iE}\Delta t_i)\}$
	N, S	$_q\phi_{G\|N,S}$	$_rD^{\phi_{R\|N,S}}$	$^{n\Delta t}{}_rD^0$				
II	E, W	$_q\phi_{G\|E,W}$	$_rD^{\phi_{R\|E,W}}$	$^{n\Delta t}{}_rD^0$	$^{n\Delta t}{}_rA^0(x_W, y_E) = \sum_{i=1}^{n}{}_q\{(\tilde{x}_{iW}\tilde{d}_{iW}\Delta t_i) + (\tilde{y}_{iE}\tilde{d}_{iE}\Delta t_i)\}$			$_rA^{\phi_A}(z_S) = \sum_{i=1}^{B}{}_q\{(\tilde{z}_{iS}\tilde{d}_{iS}\Delta t_i)\}$
	N, S	$_q\phi_{R\|N,S}$				$_rD^{n\Delta t}$	$D^{\phi_A\|E,W}$	
III	E, W	$_q\phi_{R\|E,W}$	No calculation is performed, as an amber display indicates transition period. Further, it is anticipated that the computer need not interrogate the signal until the predetermined $_q\phi_{G\|N,S_{min}}$ has elapsed.					
	N, S	$_q\phi_{A\|N,S}$						
IV	E, W	$_q\phi_{A\|E,W}$	No calculation is performed, as an amber display indicates transition period. Further, it is anticipated that the computer need not interrogate the signal until the predetermined $_q\phi_{G\|E,W_{min}}$ has elapsed.					
	N, S	$_q\phi_{R\|N,S}$						

DECISION FUNCTION

One must now consider the basis for choosing a particular action from the alternatives available. A function, $_q\Theta_{n\Delta t}$, may be developed which expresses the difference in relative delays and savings within a subnetwork for the two assumed alternatives at intersection q.

In terms of previously defined notation, the following equation is introduced:

$$_q\Theta_{n\Delta t} = {}_q\Theta_{n\Delta t} + {}^{n\Delta t}{}_qS' - {}^{n\Delta t}{}_qD' \qquad \text{(E-79)}$$

or

$$_q\Theta_{n\Delta t} = {}^{n\Delta t}{}_q\{({}_qS - {}_qD) + ({}_rS - {}_rD) + \cdots ({}_uS - {}_uD)\} \qquad \text{(E-80)}$$

The computer will perform the following test:
If

$$_q\Theta_{\Delta t} \geqq 0 \qquad \text{(E-81)}$$

extend the current phase, because more or equal delay is predicted to be accrued by terminating the current phase. However, if

$$_q\Theta_{\Delta t} < 0 \qquad \text{(E-82)}$$

less delay would be incurred by changing the phase now than Δt from now. However, it may be necessary to calculate $_q\Theta_{2\Delta t}$, $_q\Theta_{3\Delta t}$, . . ., $_q\Theta_{n\Delta t}$, because less delay may be incurred at these other times. It should be noted that if any of the subsequent $_q\Theta$'s are positive the procedure may terminate and the particular extension whose Θ yields a positive quantity is accepted.

NOMENCLATURE AND DEFINITIONS

q,r,s,t,u = the intersections that are considered to form the subnetwork, where the decision with respect to traffic signal display is to apply to q;

$_q\phi_{G|E,W}$, $_q\phi_{R|N,S}$ = the phase at intersection q which displays the green signal to the east-west legs and the red to the north-south legs, respectively;

$\Theta_{n\Delta t}$ = quantity employed in the case of the isolated intersection, in order to test for minimum delay between the alternatives of an immediate change in signal phase or a postponement of $n\Delta t$;

n = integral number of units of time, Δt, considered for the proposed extension of the current phase;

$S_{T|E,W}$ = relative savings to the east-west traffic in the case of the isolated intersection;

$^{n\Delta t}{}_qS'$ = relative savings within the subnetwork, for the considered $n\Delta t$ extension of the current phase at intersection q, excluding the savings accrued at q;

$^{n\Delta t}{}_{q,r}S$, $^{n\Delta t}{}_{q,s}S$, $^{n\Delta t}{}_{q,t}S$, $^{n\Delta t}{}_{q,u}S$ = the contributions to $^{n\Delta t}{}_qS'$ which represent the relative savings at each of the intersections neighboring q, to those vehicles released from q (during the alternatives) and seeking service at r, s, t, and u, respectively;

${}^{n\Delta t}{}_{q,r}D_S{}^{\phi_{R|E,W}}$,
${}^{n\Delta t}{}_{q,r}D_S{}^0$ = the delay accrued on the south leg of intersection r to those vehicles released from the east-west legs of intersection q and seeking service at r, for the alternatives of the immediate change of phase and extension $n\Delta t$, respectively (assuming current phase at q is ${}_q\phi_{G|E,W}$, $\phi_{R|N,S}$);

${}^{n\Delta t}{}_qS$ = relative savings within entire subnetwork;

${}^{n\Delta t}{}_qD$ = relative delay within entire subnetwork;

${}_q\tilde{x}_{\iota E}$, ${}_r\tilde{y}_{\iota W}$, ${}_s\tilde{z}_{\iota S}$ = estimates during the i^{th} interval of the proportion of left turns from the east leg of intersection q, right turns from the west leg of intersection r, and straight-throughs from the south leg of intersection s, respectively;

${}^{n\Delta t}{}_{q,r}A^0(x,y,z)$ = prediction of the number of arrivals at r resulting from the extension $n\Delta t$ of the current signal phase at q (i.e., ${}_q\phi_{G|E,W}$);

${}^1{}_{4,5}\{A+B\}_3{}^2 = {}^1{}_{4,5}A_3{}^2 + {}^1{}_{4,5}B_3{}^2$

D_{dec}, D_q, D_{acc} = components considered to contribute to the delay of each vehicle; namely, that delay attributed to the deceleration necessary to fall in behind existing queue, time spent within the queue, and finally the vehicles' acceleration to free speed;

${}_rK_{S|E,W}{}^0$ = integral number of time units, Δt, over which the delay in queue is estimated to be accrued. The subscripts indicate that this queue is formed on the south leg of intersection r, resulting from departures from the east-west legs of intersection q;

${}_{q,r}\tilde{N}_S{}^0$ = prediction of the length of queue (number of vehicles) on the south leg of intersection r at the time of the first arrival from the legs at q which receive the current green extension;

${}_r\tilde{d}_{\iota S}$ = predicted departure rate, in vehicles per second, during the i^{th} interval from the south leg of intersection r;

${}_r\tilde{a}_{\iota S}$ = predicted arrival rate, in vehicles per second, to the south leg of intersection r for the i^{th} interval;

${}_1\tilde{a}'_{\iota S}$ = predicted arrival rate defined over a particular interval and zero elsewhere;

m^0 = predicted integral number of time units, Δt, necessary to serve vehicles originally queued at r, measured from the instant that the first vehicle arriving from intersection q joins the existing queue at r until this existing queue is served. The superscript signifies that the vehicles considered are arrivals from q during the current green extension;

M^0 = predicted integral number of time units, Δt, over which departures from q arrive at r. The superscript signifies that the vehicles considered are arrivals from q during the proposed (green) extension;

$\overline{(A+B)}$
$= \overline{A} + \overline{B}$ = time average of A and B over specific interval of time;

${}_rK^{0\prime}{}_{S|E,W}$ = predicted integral number of time units, Δt, referenced from the time of the first arrival to the south leg of r, until the queue on this leg is dissipated. The superscript 0 indicates that the vehicles are those arriving from the legs at q which receive the proposed (green) phase extension (east-west legs of q);

${}_1K^{0\prime\prime}{}_{S|E,W}$ = predicted integral number of time units, Δt, referenced from the time of the first arrival to the south leg of r, during which the last vehicle being considered from the east-west (green) extension at q is serviced at r;

a = predicted integral number of time units, Δt, measured between the time of the first arrival on the south leg of r which would have resulted by extending the current phase at q; and the first arrival from the same legs at q, which would have resulted from the deferred green phase (termination of current phase);

b, b' = predicted distances the same vehicle would travel to join queue at the downstream intersection if it were released during current extension or after waiting for the new red phase, respectively;

V_{avg}, V'_{avg} = predicted average velocities over b and b', respectively;

M^{ϕ_R}, m^{ϕ_R},
$K^{\phi\prime}{}_R$, $K^{\phi\prime\prime}{}_R$ = similar to the definitions given for the respective notation with the 0 superscript; however, the ϕ_R superscript signifies that these predictions apply to the same vehicles released after waiting for the next red phase, which would have occurred if signals were changed immediately;

D_{TS}, D_{TN} = delay accrued on the north and south legs, respectively, for an isolated intersection;

${}_qD_{T|N,S}$ = delay accrued on the north and south legs of intersection q;

${}^{n\Delta t}{}_qD'$ = relative delay within the subnetwork, resulting from the proposed termination of the current phase at intersection q, excluding the savings accrued at q;

${}^{n\Delta t}{}_{q,r}D$, ${}^{n\Delta t}{}_{q,s}D$,
${}^{n\Delta t}{}_{q,t}D$, ${}^{n\Delta t}{}_{q,u}D$ = the contributions to ${}^{n\Delta t}{}_{q}D'$ which represent the relative delays at each of the intersections neighboring q, to those vehicles released from q (during the alternatives) and seeking service at r, s, t, and u, respectively;

${}_{q,r}D_{qS}{}^{n\Delta t}$,
${}_{q,r}D_{qS}{}^{\phi_{A|E,W}}$ = the delay accrued within the queue on the south leg of intersection r to those vehicles released from the north-south legs of q and seeking service at r; for the alternatives of releasing the vehicles after the proposed red extension or terminating current phase and releasing the vehicles after the amber phase, respectively;

β = predicted integral number of time units, Δt, between the time of the first arrival on the south leg of r, which would have resulted by extending the current phase at q, and the first arrival from the opposite pair (north-south) of legs resulting from the decision to change the signal phase immediately;

M^{ϕ_A} = predicted integral number of time units, Δt, over which departures from q arrive at r. The superscript signifies that the vehicles considered depart during the proposed next green phase which would be initiated if the decision was to terminate the current phase immediately;

B = predicted integral number of time units, Δt, of the duration of the new green phase (as displayed to north-south legs);

m^{ϕ_A} = predicted integral number of time units, Δt, which are necessary to serve the vehicles originally queued at r, measured from the instant that the first vehicle arriving from intersection q joins the existing queue at r until the existing queue is served. The superscript signifies that the arrivals considered are from q during the next green phase, if begun immediately ($\phi_{G|N,S}$);

${}_{r}\tilde{N}_{1S}{}^{\phi_A}$ = prediction of the length of queue (number of vehicles) on the south leg of intersection r at the time of the first arrival from q (north-south legs) during the new green phase if initiated immediately;

M^{ϕ_A}, ${}_{r}K_{S|N,S}{}^{\phi'_A}$,
${}_{r}K_{S|N,S}{}^{\phi''_A}$ = similar to the definitions given for the respective notation with the 0 superscripts; however, the ϕ_A superscript signifies that these terms apply to the vehicles released from the opposite legs (north-south) during the proposed green phase, which is initiated immediately;

$M^{n\Delta t}$, $m^{n\Delta t}$,
${}_{r}K_{S|N,S}{}^{n\Delta t}$,
${}_{r}K_{S}{}^{n\Delta t''}$ = similar to the definitions given for the respective notation with the 0 superscripts; however, the $n\Delta t$ superscript signifies that these terms apply to the vehicles released from the opposite legs (north-south) during the proposed green phase which would be initiated following the extension by $n\Delta t$ of the current phase;

γ = predicted integral number of time units, Δt, between the time of the first arrival on the south leg of r, which would have resulted by extending the current phase at q, and the first arrival from the opposite pair (north-south) of legs resulting from the same decision to extend the current phase by $n\Delta t$ (deferred $\phi_{G|N,S}$); and

${}_{q}\Theta_{n\Delta t}$ = a quantity employed, in the case of the subnetwork defined by the intersection q, to test for minimum delay between the alternatives at intersection q.

APPENDIX F

CRITERIA FOR SELECTION OF THE DIGITAL COMPUTER

A phase of the current research project is to determine the speed and size of digital computer(s) to be used in implementing the developed system of equations in real time. This appendix describes the method used in determining the requirements necessary to achieve this objective of creating an adaptive control system, and also describes the computing systems currently available that will accomplish these goals.

FORMAT OF INPUT DATA FROM SENSORS

As described in Appendix C, the data coming from each intersection are grouped into 14 bits of information relating the status of each detector or sensor and the current signal state at the intersection. Also, the intersection number is generated to help detect channel-skipping errors that would result in putting the data out of synchronism. The input data format is shown in Figure F-1.

Bits 1, 2, 3 give the current state of the signal, the various signal states being as given in Table F-1. Bits 4 and 6 give the state of stop-line detectors 1 and 3 or stop-line detectors 2 and 4, depending on the state of the signal (e.g., green N-S, detectors 1 and 3 operating; or green E-W, detectors 2 and 4 operating). Figure F-2 shows the arrangement of detectors at a general intersection.

* D_1, D_2, D_3, . . . D_8, represent the possible 8 detectors located at an intersection.

TABLE F-1

SIGNAL STATES AND CODING

SIGNAL STATE N,S	E,W	CODING OCTAL EQUIVALENT OF BINARY	BINARY BIT
R	G	5	101
R	Y	1	001
G	R	2	010
Y	R	0	000

Bits 8, 10, 12, and 14 give the condition of detectors 5, 6, 7, and 8, respectively. Bits 15 through 21 give the current intersection number of the data being displayed. (The seven bits allow for 128 intersections.)

FORMAT OF OUTPUT DATA TO INTERSECTION CONTROL SIGNAL

When the computer program decides that a particular intersection should change the state of the signal, an output word consisting of 9 bits is generated. Figure F-3 shows the output bit configuration. When bit 1 contains a 1, the hold circuit switches the signal from local control to computer control, or vice versa. When bit 2 contains a 1, the actuate circuit changes the signal state. Bits 3 through 9 contain the number identifying the intersection that is to be changed.

DATA INPUT RATE

A sampling rate of 60 samples per second is necessary, due to considerations of multilane coverage with a single loop detector as discussed in Appendix B.

To determine the presence of a car at any approach detector there must be a 0 from the previous detector reading and a 1 from the present detector reading. That is, a search must be made for a 01 combination at any detector 5 through 8 over a previous and present data reading.

To determine the departure of a vehicle at the stop-line detector, there must be a 1 from the previous detector reading and a 0 from the present detector reading.

When the approaching or departing vehicle is sensed by

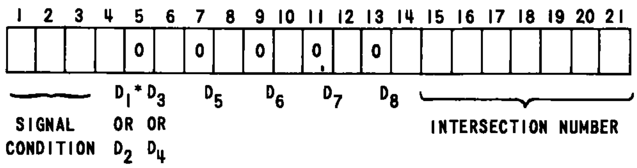


Figure F-1. Input data format.

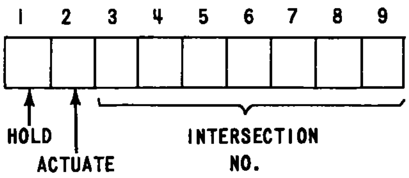


Figure F-3. Output data format.

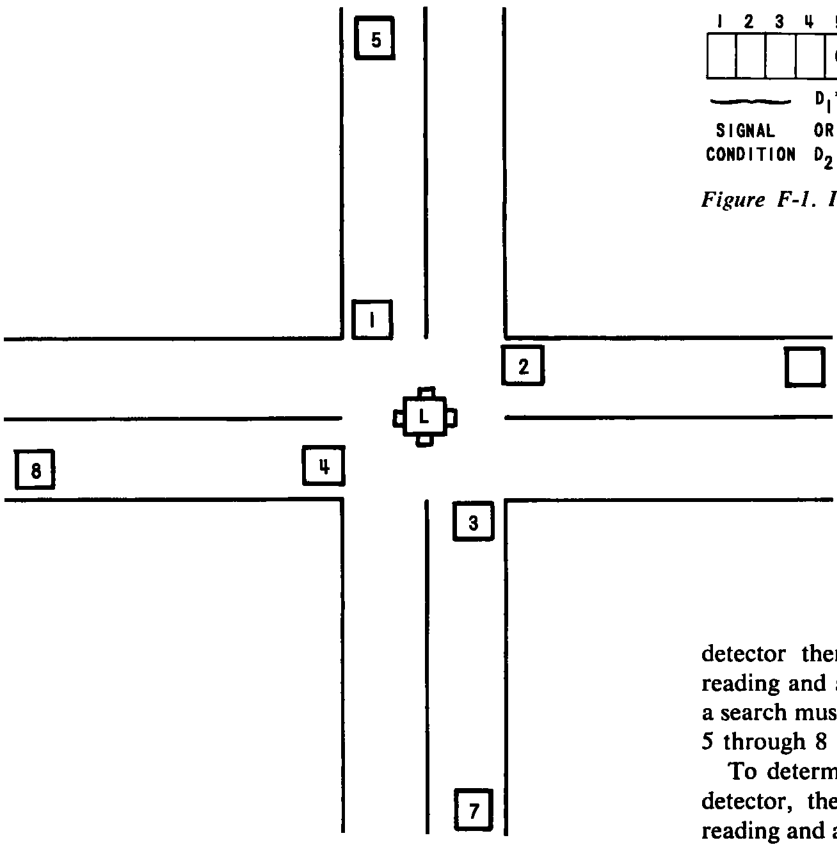


Figure F-2. Typical detector locations at an intersection.

a detector, an appropriate storage location is updated by the computer. This technique yields a current vehicle count to be employed in the programmed calculations.

DETECTOR DATA ENTRY INTO COMPUTER

The data from the detectors are presented to the computer in a sequential mode; that is, fully multiplexed (as described in Appendix C) through an I/O data channel.

DECODING METHOD USED IN COMPUTER

The data word is read into the computer through an I/O channel. Then the intersection number is extracted from this data word and placed into an index register, which allows for data to be stored in the proper place of the data set array. Masking out all but bits 7 through 14 of the data word leaves the bits for approach detectors 5 through 8. Now, the previously stored bits for this intersection are shifted left one place, the present data are added, and the sum is placed in an index register. This index register can be used to transfer control to a particular location in a list of instructions that will then transfer to the appropriate coding to update the vehicle counts. To illustrate further, this transferring technique is analogous to the "computed GO TO" statement in FORTRAN; i.e., for the 256 combinations the FORTRAN statement appears as

GO TO (1, 2, 3, . . . , 255, 256), IR

where IR contains the sum produced by shifting and adding the previous and present data together.

By shifting and adding the previous data to the present data the presence of a car can be determined at any one, all, or any combination of the four approach detectors (D5 to D8) immediately. Of the 256 combinations possible from the eight bits containing the approach detector information, 175 yield an indication of at least one car present. The possible detector combinations, together with the decimal equivalent of the binary coded (8-bit) word that will produce these combinations, are given in Table F-2. The transfer list contains 256 values, each of which represents the appropriate location to update the vehicle counters that these detectors have indicated.

To illustrate the method of decoding, the previous set of data is assumed to contain all 0's and the present set of data to contain all 1's. Then the bit configuration when combined will be as shown in Figure F-4. When these data are loaded into an index register and transferred to the combination list, they will go to location 125 octal or 85 decimal and update the vehicle counts of all four detectors.

To decode the two stop-line detectors a similar approach was taken using only 4 bits of data, giving 16 possible combination sets. Also, a similar transfer list is set up.

D_5		D_6		D_7		D_8	
0	1	0	1	0	1	0	1

Figure F-4. Bit configuration for a vehicle presence at each detector.

TABLE F-2

CODED COMBINATION OF DETECTORS

POSSIBLE DETECTOR COMBINATION	DECIMAL EQUIVALENT OF BINARY CODED (8-BIT) WORD
D8	1, 9, 13, 33, 41, 45, 49, 58, 61, 129, 137, 141, 161, 169, 173, 177, 185, 189, 193, 201, 205, 225, 233, 237, 241, 249, 253
D7	4, 6, 7, 36, 38, 39, 52, 54, 55, 132, 134, 135, 164, 166, 167, 180, 182, 183, 196, 198, 199, 228, 230, 231, 244, 246, 247
D6	16, 18, 19, 24, 26, 27, 28, 30, 31, 144, 146, 147, 152, 154, 155, 156, 158, 159, 208, 210, 211, 216, 218, 219, 220, 222, 223
D5	64, 66, 67, 72, 74, 75, 76, 78, 79, 86, 98, 99, 104, 106, 107, 108, 110, 111, 112, 114, 115, 120, 122, 123, 124, 126, 127
D7, D8	5, 37, 53, 133, 165, 181, 197, 229, 245
D6, D8	17, 25, 29, 145, 153, 157, 209, 217, 221
D6, D7	20, 22, 23, 148, 150, 151, 212, 214, 215
D6, D7, D8	21, 149, 213
D5, D8	65, 73, 77, 97, 105, 109, 113, 121, 125
D5, D7	68, 70, 71, 100, 102 103, 116, 118, 119
D5, D7, D8	69, 101, 117
D5, D6	80, 82, 83, 88, 90, 91, 92, 94, 95
D5, D6, D8	81, 89, 93
D5, D6, D7	84, 86, 87
D5, D6, D7, D8	85

TIMING REQUIREMENTS

In a worst-case analysis of the foregoing data decoding scheme—that is, where a car is detected at each detector simultaneously—it required 76 computer cycles (*8*) per intersection to decode, based on programming the research agency's current computing system.

Based on the proposed 60-samples-per-second scanning rate, 16.67 millisec are available to decode 116 intersections, or 143.6781 μsec per intersection. Therefore, the time per computer cycle is $143.6781/76 = 1.89$ μsec/cycle, and the computer must have a cycle speed greater than 1.89 μsec.

STORAGE REQUIREMENTS

The analysis has shown that 2,011 storage locations will be needed, allocated as follows:

Previous D5-D8 detector data	116
Previous D1-D3 detector data	116
Previous D2-D4 detector data	116
Vehicle counts at each detector	928
Computer program (approx.)	400
Contingency and/or expansion	335
Total	2,011

SUMMARY OF DATA ACQUISITION REQUIREMENT

From the foregoing analysis, it is found that the computer must have the following requirements:

1. A cycle speed greater than 1.89 microseconds.
2. A minimum core storage of 2,000 words (each word should have 18 bits).
3. Address modification capability (i.e., index registers or general registers).
4. A high-speed data channel (125,000 words/sec transfer rate).

GENERAL FLOW FOR IMPLEMENTATION OF MINIMUM-DELAY CRITERION

Appendices D, E, and G define the "saving" and "delay" equations as they apply to a given intersection. For implementation of these equations the general flow chart in Figure F-5 was constructed.

Generally, the sequence of operations performed by the digital computer is as follows:

1. Determine current state of signal at intersection q. Possible states for symmetrical display are as follows:

E,W	N,S
$_q\phi_{G\|E,W}$	$_q\phi_{R\|N,S}$
$_q\phi_{R\|E,W}$	$_q\phi_{G\|N,S}$
$_q\phi_{R\|E,W}$	$_q\phi_{A\|N,S}$
$_q\phi_{A\|E,W}$	$_q\phi_{R\|N,S}$

2. If the total elapsed green time (i.e., $_q\phi_{G|E,W}$, measured in seconds from the instant the phase changes to $_q\phi_{G|E,W}$; assuming that is the current signal state) is less than or equal to the minimum green time (i.e., $_q\phi_{G|E,W_{min}}$, stored for (possibly) each approach leg for each intersection) proceed to intersection q + 1. If not (i.e., greater than), go to step 3.

3. Test to determine if a vehicle (i.e., on either leg) has been waiting during the red phase, $_q\phi_{R|N,S}$. If yes, determine the time of arrival of the vehicle, $_q\phi^a{}_{R|N,S}$, measured in seconds from the time base of the change of signal to $_q\phi_{R|N,S}$. Subtract this arrival time, $_q\phi^a{}_{R|N,S}$, from the elapsed red time, $_q\phi_{R|N,S}$. If greater than the predetermined (fixed storage) $\phi_{R\,max}$, change the signal phase and proceed to intersection q + 1. If not (i.e., less than or equal to), go to step 4.

4. Each intersection, q, to be evaluated is assumed to be a member of a subnetwork of intersections of (at least one and) no more than five (q, r, s, t, and u).

5. Compute the total savings, $_qS_N$ (in vehicle-seconds) for the particular subnetwork defined by intersection q, based on the alternatives at q (i.e., change signal state immediately, or in $1\Delta t$, or $2\Delta t$, . . ., or $K\Delta t$), where

$$_qS_N = {}_{q,q}S_{i\Delta t} + {}_{q,r}S_{i\Delta t} + {}_{q,s}S_{i\Delta t} + {}_{q,t}S_{i\Delta t} + {}_{q,u}S_{i\Delta t} \qquad \text{(F-1)}$$

and the terms on the right side of the equation represent the "savings" at intersections q, r, s, t, and u, respectively, based on the alternative (i.e., now or in $i\Delta t$, where $i = 1, 2, \ldots, K$) then go to step 6.

6. Similar to step 5 above, with "savings" changed to "delay caused" and the equation changed as follows:

$$_qD_{Ni} = {}_{q,q}D_{i\Delta t} + \cdots + {}_{q,u}D_{i\Delta t} \qquad \text{(F-2)}$$

7. Calculate $\Theta_{i\Delta t}$, which is the total "savings," $_qS_{Ni}$, minus the total "delay caused," $_qD_{Ni}$.

8. If $\Theta_{i\Delta t}$ is positive (or zero), extend current signal state by $i\Delta t$ and proceed to intersection q + 1.

9. However, if $\Theta_{i\Delta t}$ is negative, and $i < K$, where K will be the stored upper limit (an integer), increment i to $i + 1$ and return to step 5; when $i = K$, go to step 10.

10. Choose least negative from computed Θ's (e.g., $\Theta_{1\Delta t}$, $\Theta_{2\Delta t}$, and $\Theta_{3\Delta t}$, where $K = 3$), calling this optimum value $\Theta_{i'\Delta t}$.

11. Signal should be changed in $i'\Delta t$ seconds. Proceed to next intersection, q + 1.

METHOD OF EVALUATING REQUIREMENTS

From an analysis of the equations to be programmed for a digital computer (Appendices D, E, G) it was found that the "savings" equations contained three basic expressions; that is,

1. The estimated delay saved by those vehicles serviced (from the two legs) during Δt, which otherwise would have to suffer the amber, red, and lost time resulting from an immediate change in phase (see Eq. G-41).

2. The delay saved which accrues to those vehicles which are served earlier in time, as a result of the previous departures (see Eqs. G-44 and G-45).

3. The additional relative saving within the subnetwork, expressed as the sum of the individual savings at each of the neighboring signalized intersections (see Eqs. G-55, G-56, and G-60).

The "delay" equations contained only those equations expressed by item 3 (compare Eqs. E-11 and E-48).

The equations described in the foregoing were programmed in FORTRAN IV* (*9, 10*) after some simplifying assumptions were applied. These assumptions were made to expedite the preliminary programming analysis. The major assumptions were as follows:

1. Certain predictor models that are updated use multivariate linear regression techniques and other predictor models use simple linear regression techniques. It was assumed here that all predictor models would use the simple linear regression technique.

2. In some cases particular upstream intersections (or sources) are not signalized. Hence, measurements would not be available for predicting arrivals at the intersection in question by the proposed predictor models. Therefore, another model must be used. It was assumed that the proposed model has the same storage and time requirements as the model used when the source conditions are not satisfied.

3. Upper limits were applied to the number of storage locations allocated for the measured and predicted departure rates for each stop line. The departures for that portion of the green phase which extends beyond these limits were assumed to be calculated from previous results and this information would be retrieved as required.

* FORTRAN IV is a problem-oriented automatic coding system designed primarily for scientific and engineering computations, and closely resembles the ordinary language of mathematics.

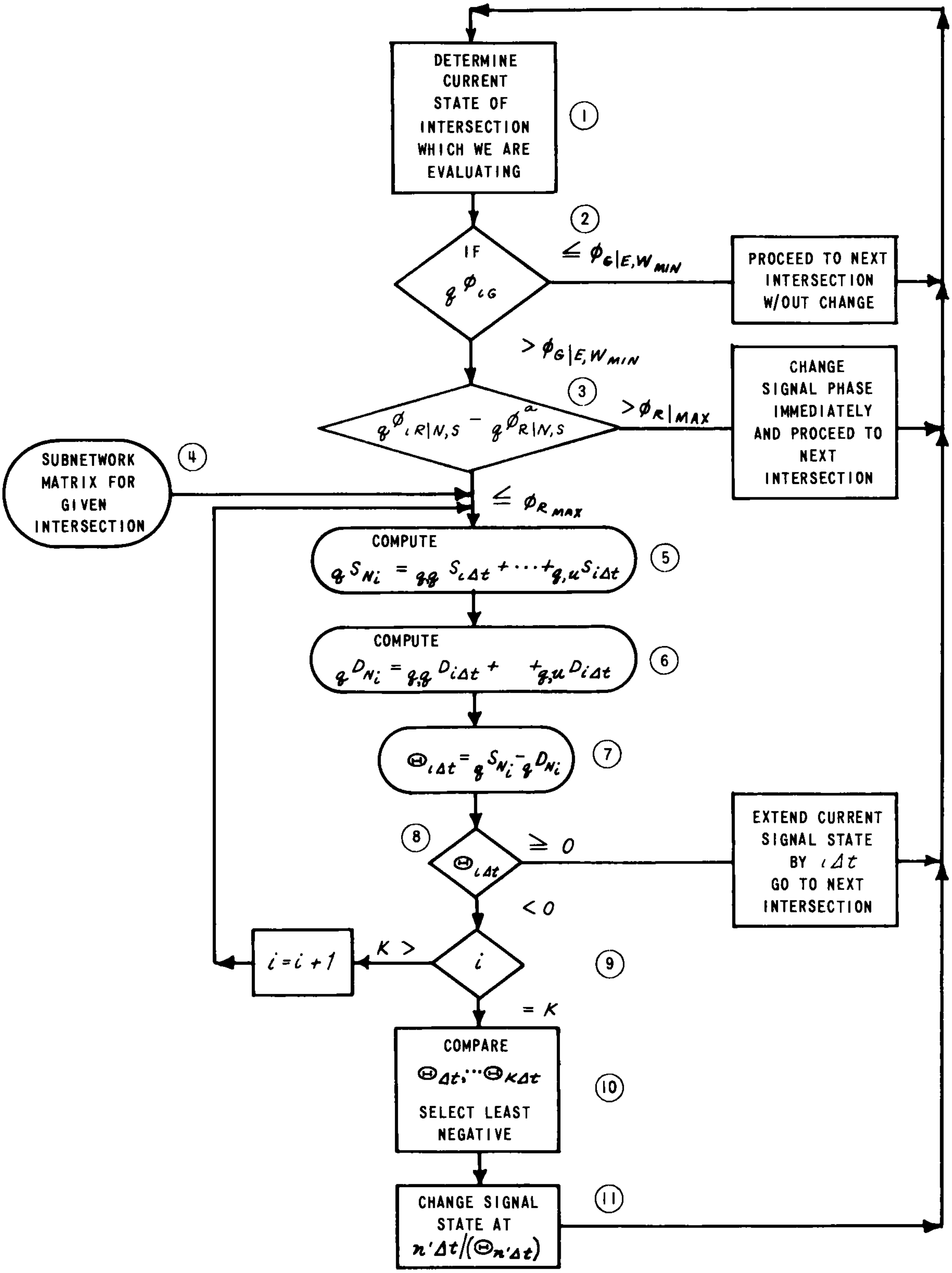


Figure F-5. General flow chart of saving and delay equation criterion.

4. Although the equations containing the arrival rates were written in terms of Δt, the arrivals were measured and predicted over periods of $2\Delta t$.

Then, using 10 cycles (*10*) for a multiplication and 18 cycles (*10*) for a division, the total number of cycles necessary was computed for each equation, with the following results:

T_1, equation Type 1, 155 cycles
T_2, equation Type 2, 565 cycles
T_3, equation Type 3, 822 cycles

The 116 signalized intersections in White Plains, N. Y., used as the typical city, were categorized into five different types of intersections. Table F-3 gives the types of intersections with the number of each type and the number of the saving and delay equations necessary at each intersection type, as well as the required number of computer cycles. The analysis shows that the total computational scheme will require 879,216 cycles, or approximately 900,000 cycles.

Also, the storage requirements for the data that will have to be stored during the calculations (e.g., measured departures, predicted departures, measured arrivals, predicted arrivals, queue lengths, length of the signal phases for each intersection), were estimated to be approximately 30,000 words of core plus 20 percent for contingencies and program instructions. Thus, approximately 37,000 words of core storage are needed.

TIMING REQUIREMENTS

From the foregoing analysis, the total calculations must be done in Δt, which has been established as 2 sec. Therefore, 900,000 cycles must be run in 2 sec, or 2.22 μsec per cycle.

SUMMARY OF COMBINED REQUIREMENTS

The analysis has given the following requirements to achieve the projected data rate and computational schemes:

1. For the computational scheme, a computer should:
   (a) Perform 900,000 cycles in 2 sec (i.e., have a computer cycle time of not more than 2.2 μsec).
   (b) Have floating point capabilities (software or hardware).
   (c) Have address modification features (i.e., index registers or general registers).
   (d) Have a minimum of 37,000 words of core storage.
   (e) Have a data channel.
   (f) Have a real-time clock.
2. For the data accumulation scheme, a computer should:
   (a) Perform 76 cycles in 143.6781 μsec (i.e., have a computer cycle time of not more than 1.89 μsec).
   (b) Contain a minimum of 2,000 words of core storage (18-bit words).
   (c) Have address modification features (e.g., index registers or general registers).
   (d) Have a high-speed data channel.

COMPUTER SYSTEMS AVAILABLE

On examining the prerequisites, it is seen that at least two possible routes are available. One method is to acquire two machines, each performing in parallel one set of the requirements (i.e., data acquisition and delay computations). Another method is to acquire one machine that can perform all requirements (i.e., both sets) sequentially in the specified time limit (2 sec).

Table F-4 details components and costs for a single-processor computing system, called System A, that will be able to fulfill all of the requirements within the specified time limits. This system will allow overlapping of the data gathering and computation functions, thus allowing the disk storage to serve as an extension of core storage. This is necessary because only 32,000 words are available and almost 39,000 are required.

Table F-5 details components and costs for a two-processor computing system, called System B, that will be able to fulfill all of the requirements if the data gathering

TABLE F-3

ANALYSIS OF SIGNALIZED INTERSECTIONS OF TYPICAL CITY [a] BY TYPE, DISTRIBUTION, AND COMPUTER CYCLES REQUIRED

INTERSECTION TYPE [b]	NO. IN TYPICAL CITY [a]	EQUATIONS REQUIRED (TYPE AND NO.) SAVING	DELAY	TOTAL CYCLES REQUIRED FOR COMPUTATION [c]
1. q, r, s, t, u	1	$T_1 + T_2 + 8T_3$	$10T_3$	15,516
2. q, r, s, t	21	$T_1 + T_2 + 6T_3$	$8T_3$	256,788
3. q, r, s	38	$T_1 + T_2 + 4T_3$	$6T_3$	339,720
4. q, r	41	$T_1 + T_2 + 2T_3$	$4T_3$	231,732
5. q only	15	$T_1 + T_2$	$2T_3$	35,460
All	116			879,216

[a] White Plains, N.Y. [b] See Figure E-2. [c] $T_1 = 155$ cycles; $T_2 = 565$ cycles; $T_3 = 822$ cycles.

TABLE F-4

COST ANALYSIS FOR A SINGLE-PROCESSOR COMPUTING SYSTEM, SYSTEM A

COMPONENT	COST ($) MONTHLY RENTAL	PURCHASE	MONTHLY MAINTENANCE FOR PURCHASE
Hardware:			
1 Processing unit with: 1-μsec cycle time/ 4 bytes[a] 131,672 bytes of core storage 16 general registers 63 instructions	6,265.00	257,700.00	250.00
consol-printing-key-board 6-line/in. pin-feed platen		59.25[b]	
channel attachment integrated single-disk storage drive, 272,000 woıds interval timer			
1 External interrupt	30.00	1,260.00	1.00
1 Floating point arithmetic (44 additional instructions)	275.00	11,550.00	11.00
1 High-speed (250 nsec[c]) general registers	700.00	29,400.00	28.00
1 High-speed (500,000 bps[d]) multiplexor channel	650.00	27,300.00	29.25
1 Low-speed (200,000 bps[d]) multiplexor channel (for card reader-punch, and printer, etc.)	350.00	14,680.00	17.75
1 Card reader punch, read 400 cpm, punch 120 cpm	525.00	26,250.00	75.50
1 Printer, 240 lines per minute 120 print positions	875.00	44,275.00	87.50
Total	9,670.00	412,424.25	500.00

Software:

Disk resident monitor with I/O capabilities
Disk resident assembler
Disk resident FORTRAN IV compiler and library
Disk resident utilities and loader
Scientific subroutine package (more than 130 FORTRAN subroutines)

[a] Byte = 8 bits. [b] This item for purchase only [c] nsec = nanosecond (10^{-9} sec). [d] bps = bytes per second.

TABLE F-5

COST ANALYSIS FOR A TWO-PROCESSOR COMPUTING SYSTEM, SYSTEM B

COMPONENT	COST ($) MONTHLY RENTAL	PURCHASE	MONTHLY MAINTENANCE FOR PURCHASE
Hardware:			
1 Central processor with: 2-μsec cycle time 4,096 words of core memory (18 bits) 8 I/O channels 8 index registers 78 instructions interval timer	2,050.00	73,800.00	250.00
1 Intercomputer coupler	600.00	24.000.00	60.00
1 Central processor with: 2-μsec cycle time integrated console with keyboard and printer 4,096 words of core memory (18 bits) 8 I/O channels 8 index registers 78 instructions interval timer	2,300.00	82,800.00	265.00
9 Additional 4,000 increments of core memory	3,150.00	113,400.00	360.00
1 I/O equipment adapter	100.00	4,000.00	15.00
1 I/O equipment 400-cpm, 80-col card reader 400-line/min printer 200-cpm, 80-col card punch	1,450.00	58,000.00	295.00
1 Code image read (binary read)	25.00	1,000.00	35.00
1 Code image punch (binary punch)	25.00	1,000.00	35.00
Total	9,700.0	358,000.00	1,315.00

Software:

FORTRAN IV compiler—load and go type, no object code produced; therefore, must recompile each time program is to run.
Assembler—"ART" (assembler for real time).
Utility package, input-output routines.
Library of scientific routines.
Floating point arithmetic.
Executive or monitor routine.

rate is relaxed to 56 samples per second (instead of 60 samples per second).

The cost analyses in Tables F-4 and F-5 indicate that both systems rent for about the same ($9,700 per month) or on a purchase-maintenance/month cost System A will be $412,424.25-$500, and System B will be $358,000-$1,315. Extrapolating over a 5½-year period both systems will cost the same—approximately $455,000, or $6,740 per month. In addition, a relatively small area must be secured and air conditioned (approximately 10 tons) to house the computing equipment. The costs associated with this section of the installation have not been included in the estimated rental, purchase-maintenance costs.

The programming time required for a program as complex as intersection control will require ½ to 1 man-year to write and debug. The major part of the programming effort will be in FORTRAN; but although FORTRAN can convert an algebraic-type statement into a machine language code, there are times when conversions do not take advantage of special features of the particular machine or it adds redundant code to achieve generality at a sacrifice of computing speed. Therefore, certain portions of the program should be written in a machine language code to achieve the efficiency necessary to a program that will run literally thousands of times per hour. The total cost for this programming effort will be approximately $25,000.

ADDITIONAL UTILIZATION OF COMPUTING SYSTEM

Aside from doing the traffic controlling, the computer system can be applied to the municipal government's commercial data processing requirements. This can be done during a part of the day when the traffic is light and the local controllers will suffice, or during the late evening hours.

The computer may be employed as a tool in the post-analysis of traffic data, as collected throughout the control periods or at other specified times, in order to improve its control function. It may also be used in the performance of accident studies, etc.

APPENDIX G

IMPLEMENTATION AND PREDICTION TECHNIQUES AS APPLIED TO THE SUBNETWORK EQUATIONS

In order to apply the philosophy as developed in Appendices D and E, a number of quantities must be predicted. The predictors and the models employed will indeed determine the sensitivity of the network's signal response to the fluctuating traffic demands. Hence, the merit of the proposed (implemented) control system will obviously rely on the judicious selection of the aforementioned prediction techniques.

One of the problems that plagues even the most advanced traffic signal system known to be operational is its reliance on certain expected values. These averages, which are used as predictions, of important traffic quantities may have the added flexibility of being varied as a function of time (clock), season (calendar), observations (human), etc. However, these predictions are generally based solely on past measurements which were observed during the last hour, day, or year. Further, the measurements may be taken at other than the particular intersection or locale to which they are applied. Such a philosophy might be acceptable for a freeway or highway where the statistics at various points may be assumed invariant as a function of location. However, applying this approach to urban traffic predictands could lead to poor control.

It is the very nature of the traffic characteristics experienced in urban areas to include large variations in the important quantities (i.e., necessary for effective, real-time control) which cannot be forecasted with "yesterday's" observations. These wide fluctuations occur not only from day to day, or hour to hour, but also within seconds. For example, consider a stop line at the north leg of a hypothetical intersection; departures have been flowing at the saturation rate since the end of the starting transient. Suddenly, a vehicle desires to turn left, but must wait for a gap in the conflicting traffic; in a matter of seconds the flow could be reduced to zero vehicles per second. One problem suggested by this example is that of obtaining a "good" prediction of the departure flow subsequent to the turning movement. Such a prediction would best be made by including all quantities which are known to be correlated with the flow rate and the most recent measurements obtained by the system. It should be noted that not all of the necessary predictions to be made are as critical or as fluctuating as the foregoing example.

The variances on the traffic quantities are believed to be attributed to both attainable and unattainable (i.e., to the system's instruments) complex factors. For example, departure rates at a particular leg of an intersection may be affected by time of day, day of week, season, weather, turning movements, actual delay experienced in the network, length of queues, special local events, international news, personal and emotional experiences, etc., some of which may be correlated with each other. Based on the foregoing discussion and the traffic data analyzed, it is believed that the system proposed should lend itself to updating techniques in order to include the most recent information, when necessary. Hence, as the computer continues to interrogate the remote instrumentation (i.e., to sample the dynamic network), the predictions would be revised to reflect this additional information. In this manner an attempt will be made to make the system respond to both sudden changes in conditions (affecting the predictions) and the more subtle variations which occur throughout the day.

Although it is not the purpose of the current study to empirically validate various prediction models, such a study would be necessary for the successful implementation of the control system, and should be performed (in part) before the system is operational. In view of this limitation, a few theoretical prediction methods which appeared to be reasonable in their applicability to the developed equations were investigated, and the empirical validation of other research work in this area was relied upon.

In selecting the prediction models the computational requirements of the digital computer were considered, because the predictand is most valuable if it can be computed simply and quickly so as to be applied within the real-time limitations. Also, this appendix makes "reasonable" simplifying assumptions, for the purpose of further reducing the computational requirements imposed on the machine by the minimum-delay equations as presented in Appendices D and E.

It is believed that good predictions of the necessary traffic quantities (i.e., those with small possible errors) may be obtained by combining simple linear regression, multiple linear regression, and average values, with a weighted moving average technique. The complexity of the particular model proposed for each of the quantities to be predicted will depend in part on the apparent physical reasonableness of the assumptions, the prediction interval required (confidence level), and the instrumentation as proposed in Appendix A. For purposes of general review a brief discussion of the applicable theory and the general equations is presented in the following.

MULTIPLE LINEAR REGRESSION

Let it be supposed that the following multiple linear regression function is to be estimated:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m + v \qquad \text{(G-1)}$$

in which

y = variable to be predicted (dependent);
v = random variable (error);
$x_1, \ldots, x_m$ = measured variables (independent); and
$\alpha, \beta_1, \ldots, \beta_m$ = unknown (to be estimated).

The expected value (unbiased) is given by

$$E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m \qquad \text{(G-2)}$$

Because the density of the random variable y is not specified, the maximum-likelihood estimators of α and the β's cannot be obtained. In these instances the least-squares method of estimation may be utilized. Since the application of these equations should yield the prediction of y at a future time from the current x values, one should collect a set of observational vectors $(y_i, x_{1i}, x_{2i}, \ldots, x_{mi})$, where y_i will be the past measured values of a quantity, which occurred at a later time than the $x_{1i}, x_{2i}, \ldots, x_{mi}$ measurements, where the sub-i indicates the time at which the observations were made and the time to which the prediction applies. It is upon this set of measurements that the estimates of α and the regression coefficients, β, in Eq. G-2 are based. In particular, the solution to the following equations will yield the desired estimates of the coefficients:

$$\begin{aligned} \text{Minimum} &= \Sigma_w v_i^2 \\ &= \Sigma_w (y_i - \alpha - \beta_1 x_{1i} - \ldots - \beta_m x_{mi})^2 \end{aligned} \qquad \text{(G-3)}$$

where

$\Sigma_w\{\ \}$ = the weighted (average) summation of $\{\ \}$ over past measurements (defined later).

$\widetilde{\{\ \}}$ = an estimate of $\{\ \}$.

Setting the partial derivatives of Eq. G-3 equal to zero gives

$$\left.\begin{aligned} &\frac{\partial}{\partial \alpha}\{\Sigma_w (y_i - \alpha - \beta_1 x_{1i} - \ldots - \beta_m x_{mi})^2\} = 0 \\ &\frac{\partial}{\partial \beta_1}\{\Sigma_w (y_i - \alpha - \beta_1 x_{1i} - \ldots - \beta_m x_{mi})^2\} = 0 \\ &\quad\vdots \\ &\frac{\partial}{\partial \beta_m}\{\Sigma_w (y_i - \alpha - \beta_1 x_{1i} - \ldots - \beta_m x_{mi})^2\} = 0 \end{aligned}\right\} \qquad \text{(G-4)}$$

whose solutions will yield the sought for estimates, $\tilde{\alpha}$, $\tilde{\beta}_1, \ldots, \tilde{\beta}_m$.

GENERAL EQUATIONS FOR ESTIMATES OF α AND β

Let, $\bar{x}_1, \ldots, \bar{x}_m$ denote the weighted average of $x_{1i}, \ldots, x_{mi}$, where

$$\left.\begin{aligned} \bar{x}_1 &= \sum_i{}_w\, x_{1i} \\ &\ \ \vdots \\ \bar{x}_m &= \sum_i{}_w\, x_{mi} \end{aligned}\right\} \qquad \text{(G-5)}$$

Further, let the a_{jk} element of the mth order determinant $|A|$ be defined as

$$a_{jk} = \sum_i{}_w\, (x_{ji} - \bar{x}_j)(x_{ki} - \bar{x}_k) \qquad \text{(G-6)}$$

where $j = 1, 2, \ldots, m,$
$k = 1, 2, \ldots, m,$

and a^*_{jk} is defined as the minors of determinant $|A|$. If p is allowed to take values of 1, 2, . . ., m, we may express the least-squares estimate $\tilde{\alpha}$ and $\tilde{\beta}_p$, of the constant term α, and the regression coefficient β_p, as follows:

$$\tilde{\beta}_p = \frac{\sum_{j=1}^{m} a^*_{pj} \sum_i{}_w\, (x_{ji} - \bar{x}_j) y_i}{|A|} \qquad \text{(G-7)}$$

$$\tilde{\alpha} = \bar{y}_i - \sum_{p=1}^{m} \tilde{\beta}_p \bar{x}_p \qquad \text{(G-8)}$$

Hence, based on the number (i.e., m) of predictors (explanatory variables) required for an individual quantity, y, one may use Eqs. G-7 and G-8 to estimate the unknowns of Eq. G-1.

USE OF AVERAGES AS PREDICTIONS

There are various types of averages which at first inspection appear to be good predictions in general. However, the characteristics required are dictated by the unique application; namely, to predict certain traffic quantities at various intersections within an urban area for use by a digital computer in making real-time control decisions. With respect to this application, the properties of these averages (and predictions in general) appear to be as presented in the following. The averaging (prediction) techniques should:

1. Include as many of the past observations as practical;
2. Weight the observations in order to give more importance to the most recent observations than to the "most past."
3. Require simple calculations in order to minimize computer size and solution time (real-time applicability).
4. Be able to include the most recent observations without complexity (i.e., easily updatable).

In view of the foregoing, an exponentially weighted moving average is proposed for use in predicting various traffic quantities. These averages have been widely used for the operation of inventory control systems. These exponentially weighted moving averages are used to extrapolate a sales time-series (forecast sales), because they can be made quickly, easily, adaptable to electronic computers, with a minimum amount of information, and to introduce current sales information (*11*). Further, Miller (*4*) presents some empirical studies which to some degree validate the application of such averages to the predictions of certain traffic quantities needed for urban signal control.

EXPONENTIALLY WEIGHTED MOVING AVERAGES

In general, the weighted average of y may be expressed as

$$\bar{y} = \frac{w_0 y_0 + w_1 y_1 + \ldots + w_k y_k}{w_0 + w_1 + \ldots + w_k} \tag{G-9}$$

or

$$\bar{y} = \frac{\sum_{i=0}^{k} w_i y_i}{\sum_{i=0}^{k} w_i} \tag{G-10}$$

where

$w_0, w_1, \ldots, w_k$ = arbitrary non-negative constants, called weights;
y_i = past observations of y; and
$i = 0, 1, 2, \ldots, k$, where y_0 is the most current observation and y_k is the most outdated observation used in Eqs. G-9 and G-10.

The exponentially weighted moving average is defined as follows:

Let $w_0, w_1, \ldots, w_k = \rho^0, \rho^1, \ldots, \rho^k$, respectively, where $0 < \rho < 1$. Referring to Eq. G-10, and letting k approach infinity (i.e., including all the measurements of the time series), the exponentially weighted moving average of y may be written as

$$\bar{y} = \frac{\sum_{i=0}^{\infty} \rho^i y_i}{\sum_{i=0}^{\infty} \rho^i} \tag{G-11}$$

which may be simplified by rewriting the denominator (infinite series) to express its limiting value; namely,

$$(1 + \rho + \rho^2 + \rho^3 + \ldots) = (1 - \rho)^{-1} \tag{G-12}$$

Substituting Eq. G-12 in Eq. G-11 gives

$$\bar{y} = (1 - \rho) \sum_{i=0}^{\infty} \rho^i y_i \tag{G-13}$$

It should be noted that $\bar{y}$ is employed as the prediction of y (namely, $\tilde{y}$) during a future time (i.e., $\bar{y} \equiv \tilde{y}$). Hence, the following may be defined, to denote the time-series of predictions:

$\tilde{y}_0 \equiv \bar{y}_0 =$ Prediction of y, which includes the observations y_0 to y_∞, inclusive; and
$\tilde{y}_1 \equiv \bar{y}_1 =$ Prediction of y, which includes the observations y_1 to y_∞, inclusive.
.
.
.

Rewriting Eq. G-13, with the objective of developing a recursive equation to be implemented by a digital machine, gives

$$\tilde{y}_0 = (1 - \rho) \left\{ y_0 + \sum_{i=1}^{\infty} \rho^i y_i \right\} \tag{G-14}$$

or

$$\tilde{y}_0 = (1 - \rho)\, y_0 + (1 - \rho) \sum_{i=1}^{\infty} \rho^i y_i \tag{G-15}$$

Examining the second term on the right side of Eq. G-15,

$$(1 - \rho) \sum_{i=1}^{\infty} \rho^i y_i = (1 - \rho) \sum_{i=1}^{\infty} \rho\, \rho^{(i-1)} y_i \tag{G-16}$$

$$= \rho \left\{ (1 - \rho) \sum_{i=1}^{\infty} \rho^{(i-1)} y_i \right\} \tag{G-17}$$

Upon further reflection, the term in brackets is merely the previous prediction, $\tilde{y}_1$, which does not include the current measurement y_0. Therefore, Eq. G-17 becomes

$$(1 - \rho) \sum_{i=1}^{\infty} \rho^i y_i = \rho\, \tilde{y}_1 \tag{G-18}$$

Finally, Eqs. G-15 or G-13 may be expressed as

$$\tilde{y}_0 = (1 - \rho)\, y_0 + \rho\, \tilde{y}_1 \tag{G-19}$$

or, in another form,

$$\tilde{y}_0 = y_0 + \rho\, (\tilde{y}_1 - y_0) \tag{G-20}$$

In words, Eq. G-20 states that in order to update the exponentially weighted moving average so as to include the most recent observation, y_0, one must simply add to this measurement the product of ρ and the difference between the measurement and its (previous) prediction, $(\tilde{y}_1 - y_0)$, where $\tilde{y}_0$ is the updated (prediction) weighted moving average of the quantity y, y_0 is the most recent measurement of the quantity y, and $\tilde{y}_1$ is the last weighted moving average of y that was employed as the prediction of the measurement y_0.

INCLUDING THE EXPONENTIALLY WEIGHTED MOVING TECHNIQUE IN THE MULTIPLE LINEAR REGRESSION MODEL

Recalling Eqs. G-5, the following define the weighted sums of $x_{1i}, \ldots, x_{mi}$:

$$\left.\begin{aligned} \bar{x}_1 &= \sum_{i=0}^{\infty} (1-\rho)\,\rho^i x_{1i} \\ &\;\vdots \\ \bar{x}_m &= \sum_{i=0}^{\infty} (1-\rho)\,\rho^i x_{mi} \end{aligned}\right\} \tag{G-21}$$

where the weights are again exponential, and Eq. G-21 is the exponentially weighted moving average of the predictor variables, $x_{1i}, \ldots, x_{mi}$.

In general, the term $\sum_{w\,i}(\;)$ included in Eqs. G-3 through G-8 can now be defined as follows:

$$\sum_{w\,i} (\;) = \sum_{i=0}^{\infty} (1-\rho)\,\rho^i (\;)_i \tag{G-22}$$

Therefore, when one seeks to update the estimates of the regression coefficients, β_p, it becomes apparent, on expanding Eq. G-7, that the inclusion of the most recent data is affected by updating various exponentially weighted moving averages. The number of terms required to be updated at any one time for a particular prediction, $\tilde{y}_i$, will depend on the number of predictors found necessary to be included in the particular regression model for y.

SOLUTIONS—$\tilde{\alpha}$, $\tilde{\beta}_p$

This section presents the expanded solutions for $\tilde{\alpha}$ and $\tilde{\beta}_p$. Two cases—namely, simple linear regression where $m = 1$ (Case 1) and multiple linear regression where $m = 2$ (Case 2)—are given.

For Case 1:

$$\tilde{y} = \tilde{\alpha} + \tilde{\beta}_1 x_1 \tag{G-23}$$

$$\tilde{\alpha} = \bar{y} - \tilde{\beta}_1 \bar{x}_1 \tag{G-24}$$

$$\tilde{\beta}_1 = \frac{\sum_{w\,i} y_i x_{1i} - \bar{x}_1 \bar{y}}{\sum_{w\,i} x_{1i}^2 - \bar{x}_1^2} \tag{G-25}$$

For Case 2:

$$\tilde{y} = \tilde{\alpha} + \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2 \tag{G-26}$$

$$\tilde{\alpha} = \bar{y} - \tilde{\beta}_1 \bar{x}_1 - \tilde{\beta}_2 \bar{x}_2 \tag{G-27}$$

$$\tilde{\beta}_1 = \frac{\sum_{w\,i} x_{1i} x_{2i} \sum_{w\,i} x_{2i} y_i - \sum_{w\,i} x_{2i}^2 \sum_{w\,i} x_{1i} y_i}{\left(\sum_{w\,i} x_{1i} x_{2i}\right)^2 - \sum_{w\,i} x_{1i}^2 \sum_{w\,i} x_{2i}^2} \tag{G-28}$$

$$\tilde{\beta}_2 = \frac{\sum_{w\,i} x_{2i} x_{1i} \sum_{w\,i} x_{1i} y_i - \sum_{w\,i} x_{1i}^2 \sum_{w\,i} x_{2i} y_i}{\left(\sum_{w\,i} x_{1i} x_{2i}\right)^2 - \sum_{w\,i} x_{1i}^2 \sum_{w\,i} x_{2i}^2} \tag{G-29}$$

Now, a partitioned-column matrix whose elements are those contained in Eqs. G-24, G-25, G-27, G-28, and G-29 (i.e., Cases 1 and 2, respectively), is denoted as follows:

$$\Psi = [\Psi^1 \mid \Psi^2]^{\mathrm{T}} \tag{G-30}$$

where Ψ, Ψ^1, and Ψ^2 are necessarily all column matrices, or, more precisely,

$$\Psi = [\underbrace{y_i\; x_{1i}\; y_i x_{1i}\; x_{1i}^2}_{\text{case 1}} \mid \underbrace{y_i\; x_{1i}\; x_{2i}\; x_{1i}x_{2i}\; x_{2i}y_i\; x_{1i}y_i\; x_{1i}^2 x_{2i}^2}_{\text{case 2}}]^{\mathrm{T}} \tag{G-31}$$

where T denotes the matrix transpose.

Now, writing the recursive equation (i.e., updating the exponentially weighted moving averages) for the general matrix Ψ^s, where s denotes the case number (i.e., $s = 1, 2, \ldots$) the following is obtained:

$$\left(\sum_{w\,i} \Psi^s\right)_{\text{UPDATED}} = \Psi^s{}_{\text{CURRENT}} + \rho \left\{\left(\sum_{w\,i} \Psi^s\right)_{\text{PREVIOUS}} - \Psi^s{}_{\text{CURRENT}}\right\} \tag{G-32}$$

In the cases previously discussed, $s = 1$ and 2. The number of individual (non-matrix) updating equations necessary would depend on the number of explanatory variables. In the aforementioned cases (1 and 2) it would require 4 and 8 (i.e., number of elements in Ψ^1 and Ψ^2) equations, respectively.

PROPOSED PREDICTION MODELS

Assuming the general subnetwork q, r, s, t, u (Fig. E-2) with the current state of signal q given as $\phi_{\mathrm{G|E,W}}$, $\phi_{\mathrm{R|N,S}}$, the following suggests particular prediction techniques for many of the quantities developed in the equations of Appendices D and E.

$\phi_{\mathrm{R|E,W}}$

The predicted duration of the next red phase (in seconds) may be obtained from

$$\{\tilde{y} = \tilde{\alpha} + \tilde{\beta}_1 x_1\}^{\phi_{\mathrm{R|E,W}}} \tag{G-33}$$

where $\tilde{y}^{\phi_{\mathrm{R|E,W}}}$ is the prediction of the duration of the next red phase (in seconds) as displayed to the east-west legs; and, $x^{\phi_{\mathrm{R|E,W}}}$ is the observed number of vehicles, queued on the opposite legs (combined, north and south) at the instant $\phi_{\mathrm{G|E,W}_{\min}}$ has elapsed.

$\phi'_{\mathrm{A|E,W}}$

The prediction of that part of the amber phase that may be considered as an extension of the red phase (i.e., during which zero flow occurs) may be obtained from the following average:

$$\{\tilde{y} = \tilde{\alpha}\}^{\phi'_{\mathrm{A|E,W}}} \tag{G-34}$$

where $\tilde{y}^{\phi'_{\mathrm{A|E,W}}}$ is the prediction of $\phi'_{\mathrm{A|E,W}}$, in seconds, and $\tilde{\alpha}^{\phi'_{\mathrm{A|E,W}}}$ is the exponentially weighted moving average of the past measurements of $\phi'_{\mathrm{A|E,W}}$. However, one should note that $\phi'_{\mathrm{A|E,W}}$ is not directly observable with the proposed instrumentation and therefore must be calculated.

Because the duration of the amber phase, $\phi'_{A|E,W}$, is known (i.e., in general its length will not be varied from cycle to cycle), the calculation of the past $\phi'_{A|E,W}$ values should be performed as the difference between the total duration of the amber phase and that portion of $\phi_{A|E,W}$ during which the stop-line detector measures a non-zero vehicular flow (i.e., departures).

$l_{E,W}$

The lost time, as defined, is attributed to the starting transient and may be predicted as follows:

$$(\tilde{y} = \tilde{a})^{l_{E \text{ or } W}} \qquad \text{(G-35)}$$

where $\tilde{y}^{l_{E \text{ or } W}}$ is the prediction of the lost time, in seconds, which occurs during the first few Δt intervals of the phase $\phi_{G|E,W}$; and $\tilde{a}^{l_{E \text{ or } W}}$ is the exponentially weighted moving average of the past l_E or l_W values.

Again, $l_{E \text{ or } W}$ is not directly observable with the proposed instrumentation and therefore must be calculated. These calculations will be performed on the past departure flow rates for the starting-transient period and the measured saturation flow rates from the same intersection.

Let $d^s{}_E$ be the measured saturation flow rate from the east leg; $\bar{d}^T{}_E$ the calculated average value of departure rates from the east leg during starting transient; and $T\Delta t$ the observed duration of transient. Then

$$l_E = \left(T\Delta t - \frac{\bar{d}^T{}_E}{d^s{}_E} T\Delta t \right) \qquad \text{(G-36)}$$

Note: If the vehicles accelerated instantaneously (i.e., $\bar{d}^T{}_E = d^s{}_E$), $l_E = 0$.

d_{tE}

Two methods for predicting flow rates are proposed. The first simply uses the exponentially weighted moving average of departure rates over corresponding intervals (referenced from the start of the green phase) of consecutive signal cycles, as follows:

$$(\tilde{y} = \tilde{a})^{d_{tE}} \qquad \text{(G-37)}$$

where $\tilde{y}^{d_{tE}}$ is the prediction of the departure rate, for a particular interval of the subsequent cycle; and $\tilde{a}^{d_{tE}}$ is the exponentially weighted moving average of the past departure rates for consecutive cycles. It should be noted that the summation (for the foregoing average) as presented in Eq. G-13 (namely, $\sum_{i=0}^{\infty}$) has the interpretation that i assumes all the values of past cycles.

The aforementioned prediction of d_E has the advantage of being simple and also yielding values which apply a full cycle in the future. However, it is believed that better predictions of d_E may be necessary in certain applications to the equations (i.e., Eq. D-8). Hence, the following multiple linear regression, which linearly relates both the number of vehicles in queue and the measured departure rate (during the previous time interval) to the prediction of the departure rate (during the subsequent intervals), is proposed:

$$(\tilde{y} = \tilde{a} + \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2)^{d_E} \qquad \text{(G-38)}$$

where $\tilde{y}^{d_E}$ is the prediction of the departure rate for a particular interval of the current phase; $x_1{}^{d_E}$ is the observation of the number of vehicles in queue at the time the prediction is made; and $x_2{}^{d_E}$ is the observed departure rate (current). It is believed that the general form of this prediction model would be necessary at intersections where there are left turns with the associated conflicts.

a_E

Inasmuch as arrivals at an urban intersection are usually the direct result of some control decisions (i.e., traffic signal displays), it may be possible to relate the arrivals at one intersection to the control decisions and departures of another (*4*). It should be generally noted that the a_{tE} terms considered in Appendix D and Appendix E are those arrivals to the end of the existing queue or, in the case of zero queue, the stop line. Therefore, because the predictions obtained in the following by the proposed technique are made at the approach detector, they must be extrapolated to the end of the "observed" queue in order to be applicable to the minimum-delay equations. The arrivals at point 1 (e.g., to the east leg of intersection q) of Figure G-1 are seen to be the result of straight-through movements from the east leg of intersection s during ${}_s\phi_{G|E,W}$ and the left- and right-turning movements from the south and north legs of intersection s, respectively, during phase ${}_s\phi_{G|N,S}$.

The predicted arrivals to point 1 are proposed as the following linear regression:

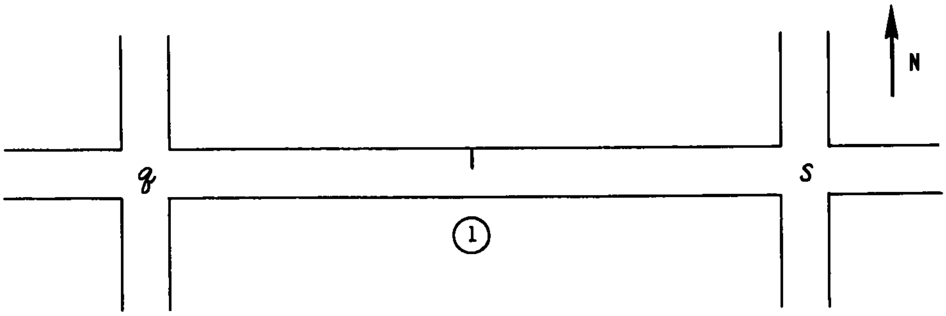


Figure G-1. Arrivals to q from s.

$$(\tilde{y} = \tilde{\alpha} + \tilde{\beta}_1 x_1)^{a_E} \qquad \text{(G-39)}$$

where

$\tilde{y}^{a_E}$ = prediction of the arrivals at point 1, for a subsequent cycle, during a particular number of intervals after the start of phase ${}_s\phi_{G|E,W}$ (or ${}_s\phi_{G|N,S}$); and

$x_1{}^{a_E}$ = total number of vehicles which desired service at q during the most current phase ${}_s\phi_{G|E,W}$ (or ${}_s\phi_{G|N,S}$).

It should be noted that $x_1{}^{a_E}$ may be calculated by summing the counts as obtained by the detector at point 1 for consecutive periods. Alternately they would apply to the results of ${}_s\phi_{G|E,W}$ and ${}_s\phi_{G|N,S}$.

SOME SIMPLIFICATIONS OF THE DERIVED EQUATIONS

Before proceeding to discuss certain predictions within the context of the individual equations, it is in order to further reduce the form of the equations, through both algebraic manipulations and reasonable simplifying assumptions. Also, a few equations as presented in Appendix D are generalized to consider an extension of $n\Delta t$.

Isolated Intersection (Appendix D)

Generalizing Eq. D-2 for an extension $n\Delta t$ gives

$${}_A S_D = {}_A\left\{\sum_{i=1}^{n}(\tilde{d}_{iE}\,\Delta t_i + \tilde{d}_{iW}\,\Delta t_i)\right\}(\tilde{\phi}_{R|E,W} + \tilde{\phi}'_{A|E,W} + \tilde{l}_{E,W}) \qquad \text{(G-40)}$$

which, through the use of average values of $\tilde{d}_i$ (denoted by $\bar{d}$) over the period $n\Delta t$, simplifies to

$${}_A S_D = \{n\Delta t\,(\bar{d}_E + \bar{d}_W)\}(\tilde{\phi}_{R|E,W} + \tilde{\phi}'_{A|E,W} + \tilde{l}_{E,W}) \qquad \text{(G-41)}$$

Generalizing Eqs. D-4 and D-5 gives

$${}_A S_{qE} = \left\{\left(\sum_{i=1}^{L}\tilde{a}_{iE}\,\Delta t_i\right) + N_{rE}\right\}\frac{\sum_{i=1}^{n} d_{iE}\,\Delta t_i}{\tilde{d}_L{}^{+}{}_E} \qquad \text{(G-42)}$$

$${}_A S_{qW} = \left\{\left(\sum_{i=1}^{L}\tilde{a}_{iW}\,\Delta t_i\right) + N_{rW}\right\}\frac{\sum_{i=1}^{n} d_{iW}\,\Delta t_i}{\tilde{d}_L{}^{+}{}_W} \qquad \text{(G-43)}$$

Reducing the form of Eqs. G-42 and G-43 by employing the average values of a_i over the period $L\Delta t$ and d_i over the period $n\Delta t$ ($\bar{a}$ and $\bar{d}$, respectively) gives

$${}_A S_{qE} = {}_A\{L\bar{a}_E\Delta t + N_{rE}\}\,\frac{(n\Delta t\,\bar{d}_E)}{\bar{d}_L{}^{+}{}_E} \qquad \text{(G-44)}$$

$${}_A S_{qW} = \{L\bar{a}_W\Delta t + N_{rW}\}\,\frac{(n\Delta t\,\bar{d}_W)}{\bar{d}_L{}^{+}{}_W} \qquad \text{(G-45)}$$

Rewriting Eqs. D-10 and D-11 and simplifying yields

$$N_{IS} + \sum_{i=1}^{k_S} a_{iS}\Delta t_i - \sum_{i=1}^{k_S} d_{iS}\Delta t_i \leq 0 \qquad \text{(G-46)}$$

$$N_{IS} + \sum_{i=1}^{k'_S}\tilde{a}_{iS}\Delta t_i - \sum_{i=1+n}^{k'_S}\tilde{d}_{iS}\Delta t_i \leq 0 \qquad \text{(G-47)}$$

or

$$N_{IS} + k_S(\bar{a}_S - \bar{d}_S)\Delta t \leq 0 \qquad \text{(G-48)}$$

$$N_{IS} + \{k'_S\bar{a}_S - (k'_S - n)\bar{d}_S\}\Delta t \leq 0 \qquad \text{(G-49)}$$

Setting Eqs. G-48 and G-49 equal to zero, and solving for k_S and k'_S, respectively, gives

$$k_S\Delta t = \frac{-N_{IS}}{(\bar{a}_S - \bar{d}_S)} \qquad \text{(G-50)}$$

and

$$k'_S\Delta t = \frac{-(N_{IS} + n\Delta t\,\bar{d}_S)}{(\bar{a}_S - \bar{d}_S)} \qquad \text{(G-51)}$$

As a result, Eq. D-12 becomes

$$\Delta t(k'_S - k_S) = \Delta t_{qS} = \frac{-n\Delta t\,\bar{d}_S}{(\bar{a}_S - \bar{d}_S)} \qquad \text{(G-52)}$$

Similar equations may be obtained for the north leg (i.e., k_N, k'_N, and Δt_{qN}).

Rewriting Eqs. D-14 and D-15, using the foregoing, gives

$$D_{\text{dec }S} = K_1(k'_S - k_S)\bar{a}_S\Delta t \qquad \text{(G-53)}$$

Substitution of Eq. G-52 in Eq. G-53 gives

$$D_{\text{dec }S} = K_1\,\frac{(-n\Delta t\,\bar{d}_S)}{(\bar{a}_S - \bar{d}_S)}\,\bar{a}_S \qquad \text{(G-54)}$$

Similarly,

$$D_{\text{acc }S} = K_2\,\frac{(-n\Delta t\,\bar{d}_S)}{(\bar{a}_S - \bar{d}_S)}\,\bar{a}_S \qquad \text{(G-55)}$$

Analogous equations may be written for the north leg (i.e., $D_{\text{dec }N}$ and $D_{\text{acc }N}$).

Subnetwork of Signalized Intersections (Appendix E)

Rewriting Eq. E-11a gives

$$\begin{aligned}
{}^{n\Delta t}{}_{q,r}D^0{}_{q,s} = {} & \sum_{j=1}^{K^0}\tilde{N}_{1S_j}\Delta t_j + \sum_{j=1}^{K^0}\left\{\sum_{i=1}^{j}(\tilde{a}'_i - \tilde{d}_i)_S\Delta t_i\Delta t_j\right\} \\
& - \sum_{j=1}^{K^0}(\bar{a}_j - \bar{d}_j)_S\frac{\Delta t_j{}^2}{2} - \sum_{j=1}^{m^0}\tilde{N}_{1S_j}\Delta t_j + \\
& \sum_{j=1}^{m^0}\left\{\sum_{i=1}^{j}\tilde{d}_{iS}\Delta t_i\Delta t_j\right\} - \sum_{j=1}^{m^0}\tilde{d}_j\frac{\Delta t_j{}^2}{2}
\end{aligned} \qquad \text{(G-56)}$$

By using the relationships that N_{1S_j} is invariant over values of $j = 1, 2, 3 \ldots K^0$, $\Delta t_i = \Delta t_j$ for any values of $i = 1, 2 \ldots$ and $j = 1, 2 \ldots$, and the simplifying assumption that the arrivals occurring over the interval $M^0\Delta t$ (at intersection r) may be included as their average over the interval $K^0\Delta t$, the following expanded form of Eq. G-56 is obtained:

$$\begin{aligned}
{}^{n\Delta t}{}_{q,r}D^0{}_{q,s} = {} & \tilde{N}_{1S}K^0\Delta t + (1 + 2 + \ldots + K^0) \\
& \times(\bar{a}_S - \bar{d}_S)\Delta t^2 - K^0(\bar{a}_S - \bar{d}_S)\frac{\Delta t^2}{2} \\
& - m^0\tilde{N}_{1S}\Delta t + (1 + 2 + \ldots + m^0) \\
& \times\bar{d}_S\Delta t^2 - m^0\bar{d}_S\frac{\Delta t^2}{2}
\end{aligned} \qquad \text{(G-57)}$$

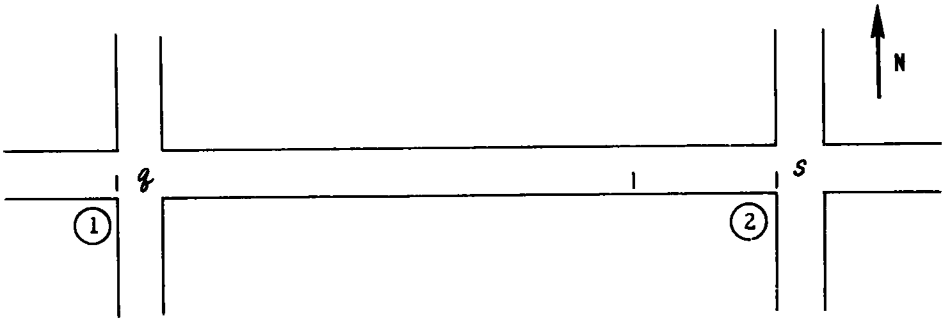


Figure G-2. Intersection pair q,s.

But

$$1 + 2 + 3 + \ldots + K = \frac{K(K+1)}{2} \qquad \text{(G-58)}$$

so Eq. G-57 may now be rewritten (rearranging and combining terms) as

$$D^0_{q,s} = N_{1S}(K^0 - m^0)\Delta t + \left\{\frac{(K^0)^2}{2}\,\bar{a}_S + \bar{d}_S\,\frac{[(m^0)^2 - (K^0)^2]}{2}\right\}\Delta t^2 \qquad \text{(G-59)}$$

which is dependent on the average values of d_S and a_S being unchanged over various time intervals (e.g., $K^0\Delta t$, $m^0\Delta t$).

Although Eq. G-59 is a simplified form of Eq. G-56, the general form is applicable to the D^{ϕ_B}, D^{ϕ_A}, and $D^{n\Delta t}$ delay terms.

CRITERIA FOR EVALUATION OF M

The equations developed in Appendix E allowed for the departures released during $n\Delta t$ from an intersection to arrive during some different time interval $M\Delta t$ at the downstream intersection. The M^0 term is first introduced in Eq. E-12 to allow for the fluctuating queue length at the downstream intersection.

It is proposed to release vehicles from the west leg of intersection q (Fig. G-2) during the extension of ${}_q\phi_{G|E,W}$ by $n\Delta t$, where

${}_S N_{OW}$ = observed queue (number of vehicles) on the west leg of s at time of the start of extension $n\Delta t$;

${}_S\bar{a}_W$, ${}_S\bar{d}_W$ = average values of the predictions of the arrivals to and the departures from the west leg of intersection s;

b_{1-2} = distance, in feet, between detectors 1 and 2;

L = average length of vehicle plus average headway (ft/vehicle);

$\bar{v}$ = average velocity over distance from detector 1 to end of queue;

$n\Delta t$ = time for first departing vehicle to reach end of queue; and

$n'\Delta t$ = time for last departing vehicle to reach end of queue.

The time at which the vehicle released from intersection q arrives at the tail end of the queue on the west leg of intersection s may be obtained from

$$\bar{v}n\Delta t = b_{1-2} - L\{{}_S N_{OW} + (\bar{a}_W - \bar{d}_W)n\Delta t\} \qquad \text{(G-60)}$$

which rearranges to

$$n\Delta t = \frac{b_{1-2} - L_S N_{OW}}{\bar{v} + L_S(\bar{a}_W - \bar{d}_W)} \qquad \text{(G-61)}$$

For the last vehicle from q,

$$\bar{v}n'\Delta t = b_{1-2} - L\{{}_S N_{OW} + {}_S(\bar{a}_W - \bar{d}_W)(n + n')\Delta t\} \qquad \text{(G-62)}$$

which simplifies to

$$n'\Delta t = \frac{b_{1-2} - L\{{}_S N_{OW} + (\bar{a}_W - \bar{d}_W)n\Delta t\}}{\bar{v} + L(\bar{a}_W - \bar{d}_W)} \qquad \text{(G-63)}$$

One may now write

$$M\Delta t = \{n + (n' - n)\}\Delta t \qquad \text{(G-64)}$$

Hence, if

$$n' > n,\ M > n \qquad \text{(G-65)}$$

or, if

$$n' < n,\ M < n \qquad \text{(G-66)}$$

However, if

$$n' = n,\ M = n \qquad \text{(G-67)}$$

Combining Eqs. G-61 and G-63 gives

$$(n' - n)\Delta t = \frac{b_{1-2} - L_S\{N_{OW} + (\bar{a}_W - \bar{d}_W)n\Delta t\} - b_{1-2}L_S N_{OW}}{\bar{v} + L_S(\bar{a}_W - \bar{d}_W)} \qquad \text{(G-68)}$$

therefore,

$$(n' - n)\Delta t = \frac{-L_S(\bar{a}_W - \bar{d}_W)n\Delta t}{\bar{v} + L_S(\bar{a}_W - \bar{d}_W)} \qquad \text{(G-69)}$$

from which one may make the following statement:

If the average rate of change of the (downstream) queue length is small with respect to the average approach velocity (for small n), one may make the approximation that $M \approx n$. Or, more concisely, if

$$\bar{v} \gg L_S(\bar{a}_W - \bar{d}_W),\ (n' - n) \approx 0 \quad \text{for small } n \qquad \text{(G-70)}$$

APPLICATION OF PROPOSED PREDICTION

Figure G-3 represents a time series of signal-cycle lengths of two faces (out of four) of a signal at intersection q, where ${}_q y_i{}^{\phi_R}$, ${}_q y_i{}^{\phi_G}$, ${}_q y_i{}^{\phi_A}$ represent measured phase lengths of ϕ_R, ϕ_G, and ϕ_A, respectively, for the ith cycle, and $i = (0, 1, 2, \ldots)$. It should be noted that with this convention, y_{i+1} occurs earlier than y_i (i.e., y_0 is the most current observation). Further, the skewed dashed lines indicate the order of continuous time (e.g., immediately upon the termination of phase ${}_q y_1{}^{\phi_A}$, ${}_q y_0{}^{\phi_R}$ is initiated).

Hence, to obtain $\tilde{y}_0{}^{\phi_R}$ (i.e., the predicted duration of the subsequent red phase), Eq. G-33 indicates that the value of the combined queue on the opposing legs, $x_{10}{}^{\phi_R}$, at time $y_0{}^{\phi_{G\min}}$ (occurring within the interval $y_0{}^{\phi_G}$) be employed as a predictor. The updating of the regression coefficients may be performed, at most, once per signal cycle. Similarly, one may also predict the phase length of the next red phase for the other pair of faces from the series in Figure G-3. Assuming that the cycle series is given for the east-west faces, it may be used to obtain $\tilde{y}^{\phi_{R|N,S}}$ in view of the following indentities:

$$y_i{}^{\phi_{R|N,S}} = y_i{}^{\phi_{G|E,W}} + y_i{}^{\phi_{A|E,W}} \tag{G-71}$$

$$y_i{}^{\phi_{G|N,S}} = y_i{}^{\phi_{R|E\,W}} - y_i{}^{\phi_{A|N,S}} \tag{G-72}$$

The prediction, $\tilde{y}_0{}^{\phi_{R|N,S}}$, may be made at the time $y^{\phi_{G|N,S\,\min}}$ after the initiation of $y^{\phi_{R|E,W}}$, and in this case includes the combined queue length on the east-west legs as the predictor. The prediction, $\tilde{y}^{\phi'_{A|E\,W}}$ is an average of all the past intervals during which departures from the east-west legs have been zero (stop-line detector observations), within the interval $y^{\phi_{A|E,W}}$. Similarly, the prediction, $\tilde{y}^{\phi'_{A|N,S}}$ may be performed by observing the stop-line detector on the north-south legs during the interval $y^{\phi_{A|N,S}}$ prior to the initialization of $y^{\phi_{G|E,W}}$. These predictions may be updated at most once per cycle.

Again, referring to Figure G-3, Eq. 37 averages departure rates for a particular interval of the east-west green phase {i.e., $1\Delta t$, $2\Delta t$ . . . or $(y^{\phi_{G|E,W}} + y^{\phi_{A|E,W}} - y^{\phi'_{A|E,W}})$} over all past $\phi_{G|E,W}$ phases (i.e., cycles 1, 2, . . .) to predict the departure rates y^{dE} or y^{dW} during the corresponding interval for the subsequent $\phi_{G|E,W}$ phase. Similarly, the predictions $\tilde{y}^{dN}$ and $\tilde{y}^{dS}$ are obtained from measurements of the departure rates for the intervals $1\Delta t$, $2\Delta t$, . . . $(y^{\phi_{R|E,W}} - y^{\phi_{A|N,S}} + y^{\phi_{A|N,S}})$, respectively, over all past $\phi_{G|N,S}$ phases to obtain the prediction for the corresponding interval for the subsequent $\phi_{G|N,S}$ phase.

NOMENCLATURE AND DEFINITIONS

y = variable to be predicted;
v = random variable;
$x_1, \ldots, x_m$ = predictor variables;
$a, \beta_1, \ldots, \beta_m$ = theoretical regression coefficients;
E ( ) = expected value of ( );
$y_i, x_{1i}, \ldots, x_{mi}$ = past measurements of the y and x quantities, respectively;
$\sum_{i}{}_{w}$ ( ) = the (arbitrary) weighted sum of ( ), which is taken over the past observations;
$(\tilde{\ })$ = an estimate of ( ) or a prediction of ( ) (depends on the particular context in which it appears);
$\frac{\partial}{\partial a}\{\ \}, \frac{\partial}{\partial \beta_1}\{\ \}, \ldots, \frac{\partial}{\partial \beta_m}\{\ \}$ = the partial derivatives of the quantity { } with respect to the regression coefficients $a, \beta_1, \ldots, \beta_m$, respectively;
$\tilde{a}, \tilde{\beta}_1, \ldots, \tilde{\beta}_m$ = estimates of the theoretical regression coefficients $a, \beta_1, \ldots, \beta_m$, respectively;
$(\bar{\ })$ = weighted average of ( ) or, in particular, the exponentially weighted moving average of ( ) or the arithmetic mean of ( ) (the definition depends on the particular context in which it appears);
$\bar{x}_1, \ldots, \bar{x}_m$ = averages over i of $x_{1i}, \ldots, x_{mi}$, respectively;
m = total number of predictors used in a function;
A = m^{th} order matrix;
$|A|$ = determinant of a;
a_{jk} = elements of A equal to $\sum_{i}{}_{w} (x_{ji} - \bar{x}_j)(x_{ki} - \bar{x}_k)$ where $j = 1, 2, \ldots, m$ and $k = 1, 2, \ldots, m$;
a^*_{jk} = minors of the determinant $|A|$;
$w_0, w_1, \ldots, w_k$ = arbitrary non-negative constants (called weights);
y_i = past observations of y (note that y_{i+1} occurs previous to y_i);
ρ = a number between 0 and +1;
ρ_i = weight of the measurement y_i;
y_0 = the most recent observation of y;

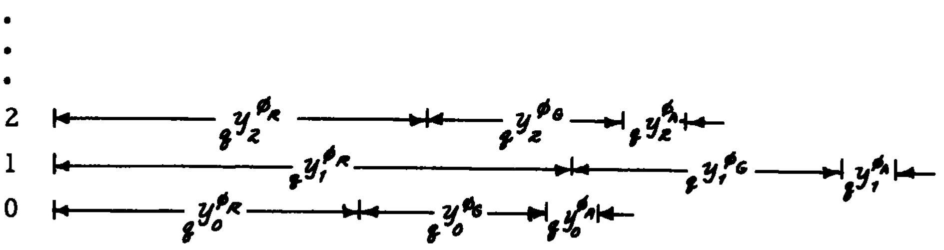


Figure G-3. Time series of signal cycles at intersection q.

$\tilde{y}_1$ or $\bar{y}_1$ = a prediction of y which includes the observations y_0 to y_x;

$\tilde{y}_1$ or $\bar{y}_0$ = the last prediction of y, which where referenced from the current instant includes the observations y_1 to y_x;

Ψ^S = a matrix whose elements are required to be updated in order to update the estimates of $a, \beta_1, \ldots, \beta_m$;

$\{A + B\}^\phi = A^\phi + B^\phi$;

$\tilde{y}^{\phi_{R|E,W}}$ = equivalent to $\tilde{\phi}_{R|E,W}$;

$x_1^{\phi_{R|E,W}}$ = observed number of vehicles queued on the opposite legs (combined north and south) at the instant $\phi_{G|E,W\,min}$ has elapsed;

$\tilde{y}^{\phi'_{A|E,W}}, \tilde{y}^{l_E}, \tilde{y}^{d_E}, \tilde{y}^{a_E}$ = equivalent to $\tilde{\phi}'_{A|E,W}$, $\tilde{l}_E$, $\tilde{d}_E$, and $\tilde{a}_E$, respectively;

$a^{\phi'_{A|E,W}}$ = an average of past measurements of $\phi'_{A|E,W}$;

$\tilde{a}^{l_E}$ = an average of the past measurements of l_E;

d^{s_E} = measured saturation flow rate from east leg;

$\bar{d}^{T_E}$ = calculated average departure rate during starting transient (east leg);

$T\Delta t$ = observed duration of transient;

b_{1-2} = distance between detectors 1 and 2;

$\bar{L}$ = average length of vehicle plus average headway;

$\bar{v}$ = average velocity;

$n\Delta t$ = time for first departing vehicle to reach end of the queue downstream; and

$n'\Delta t$ = time for last departing vehicle to reach end of the queue downstream.

APPENDIX H

APPLICATIONS OF THE HIGH-SPEED DIGITAL COMPUTER IN CONTROL OF TRAFFIC

Because the tasks that may be performed by a digital computer are in general limited only by the ingenuity of the individual developing the program logic, it is not surprising to find that the general purpose high-speed digital computer may be employed in the control of traffic at various levels. For example, the computer may be used to find solutions to particular network configurations as a function of various traffic parameters by modeling the traffic and control strategies for laboratory experimentation. This technique is commonly described as a simulation of the traffic network and its related control logic.

Although most of the work in this field has been concerned with the "degenerate network" or the isolated intersection, there have been some simulation studies for a number of signalized intersections. With the currently available computer hardware and software, it is evident that the computational complexity of the problem is formidable for a large network. Nevertheless, it is conceivable that, given any physical realizable objective function (i.e., minimizing delay, minimizing queue lengths, etc.), an optimum solution may be found through computer simulation.

Because these solutions yielding cycle length, splits, offsets, etc., vary as a function of the values of the traffic parameters, the control of the real network of signals must depend on a knowledge of these parameter values or some other related function. Further, it should be noted that limitations which are inherent within the modeling of the vehicular dynamics, human reaction, control logic, etc. (when compared to the real situation) will, in part, determine the validity of these solutions when applied to the given network. It is apparent that various degrees of realism in the building of a model are required for obtaining solutions to different control systems, networks, and constraints. These degrees of realism will depend on the available knowledge pertaining to the real situation, the approximations and/or assumptions permitted, and the philosophy used in the development of the model (e.g., macroscopic vs microscopic representation of traffic).

The computer also presents itself as an excellent tool to be used in directly controlling actual traffic signals in real time. In this function the computer would generally perform calculations to determine the value of the control parameters, decode information as presented to it by the vehicle detectors, and transmit the coded solution through the communications network to actuate the local signal controllers. In the hypothetical direct application of simulation results of an intersection controlled by a fixed signal cycle, the control computer may, instead of its calculation role, serve as an accounting device. This device has the capacity to memorize large quantities of previously obtained solutions and change the cycle length and split according to the current traffic information presented to it.

In its function as a control unit, the computer may be programmed to emulate the operation of any number of special purpose traffic controllers (full traffic-actuated, progressive, etc.) Without extending the emulation concept further, advantages may be gained merely through the flexibility of being able to apply various controls to subnetworks that may be redefined in terms of configuration and intersection content as the traffic flow warrants. In addition, the input constants (minimum green, maximum red, offsets, splits) that are normally entered mechanically at the local controller may now be varied from the control site throughout the day as conditions necessitate. Obsolescence, as experienced with the purchase of special purpose equipment through incapabilities of the selected control system to meet changes in ordinances, new signal phases, road construction, or simply the changing vehicular flow* patterns, is virtually eliminated.

It is apparent that this technique of emulating existing traffic control equipment is not using the computer optimally, because the very nature of the control philosophies employed is limited by the mechanical-electrical equipment selected. To utilize the computer's capabilities to efficiently control the traffic signals within a network, it is suggested that once a criterion, or set of criteria, has been found it should be implemented as a general set of equations that account for information obtained on the dynamic traffic characteristics. Hence, by optimizing the objective function within the constraint equations, one may develop an adaptive control system. It should be noted that simulation techniques can prove to be a powerful tool in both the "emulation control" and the more general adaptive system. By building a model to simulate the particular control logic, detector locations, boundary values, etc., may be investigated.

In further applications, as in using the computer as a data processing tool for post-analysis studies, large quantities of real traffic data obtained by automatic data gathering equipment or human observers may be analyzed quickly and efficiently. The results of such analyses are particularly important if the "figure-of-merit" of the particular control system (e.g., fixed-time or adaptive control) is not calculated as part of the controlling process.

As a result, computers may (a) be employed in obtaining a solution *a priori* for a specific network for a set of traffic parameters, (b) aid in the implementation of previously determined solutions, or (c) implement any other form of control (which is now in existence or yet to be developed), and (d) process large amounts of traffic data in order to reveal *a posteriori* an existing system's performance. However, it must be reemphasized that the validity of the results obtained from any simulation, control, or data analysis operation obviously will not be proportional to the capabilities of the machine. Rather, it will depend on the validity of the concepts and theories put forth by the individuals using the tool.

This appendix defines a number of conventional control strategies (*18, 19, 20, 21, 22*), then presents the control logic in the form of a flow chart which may be programmed for the digital machine. Also included is a discussion of more academic interest concerning computer simulation (*7, 17*) and its direct application to control.

SIMULATION

Simulation techniques are capable of providing the traffic analyst with the necessary number of traffic hours at predetermined conditions in order to obtain good estimates of the mean values of the variables of interest. Also, they are able to provide for rapid and easy changes in the control logic and traffic conditions within the program.

For purposes of illustration, the simulation of an orthogonal intersection of two, two-lane, two-way streets (Fig. H-1) is examined to determine how the results may be applied to the control of real traffic signals. In the simplest case the signal is assumed to be of the fixed-time type (predetermined cycle length, split, and amber). The simulation runs would be made in order to find the optimum solution for a given set of traffic parameters. As the traffic conditions vary the values of these traffic parameters will, in general, yield different solutions. Figure H-2 shows typical traffic parameters that may be considered in determining the most effective signal setting. Each of the four approach legs has associated with it four levels of approach volumes and two levels of turning percentages per movement. The possible permutations of the parameters of the four legs taken at one time involve 65,536 distinct cases requiring solutions. The number of traffic hours required to obtain a solution will depend on the number of hours required per control setting (to obtain the mean value of the objective function) and the number of individual control combinations considered to be feasible solutions. By multiplying the number of traffic hours required by the simulation computer time per traffic hour one may obtain the total amount of computer time necessary to obtain the solutions. The simulation time (computer time per traffic hour) will be a function of the complexity of the model, the programming techniques employed, and the characteristics of the computer used.

As an example, one "fairly complex model" (*7, 17*) run on an IBM 7090 computer to simulate conditions at a pretimed isolated intersection is examined. Only 0.5 sec of computer time was required to simulate 1 hr of traffic time, which gives a simulation time (computer time/traffic time) of approximately 1:7,200. Hence, if 144,000 traffic-hours were required for the set of solutions, approximately 20 hr of computer time would suffice. By taking actual

* Caution should be observed in the interpretation of this sentence Obviously, when appreciable increases in the efficiency of the flow of traffic through control have brought that flow to saturation, only construction of new highways and/or restriction of vehicular use will solve the traffic problem.

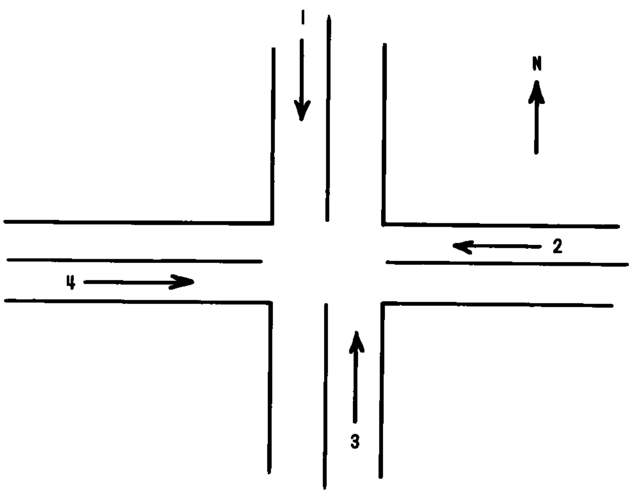
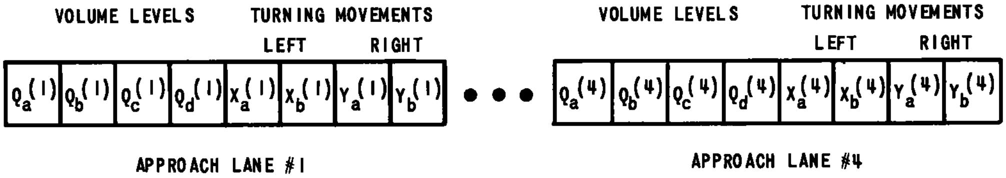


Figure H-1. Orthogonal signalized intersection of two 2-lane, two-way streets.



Q_a APPLIES IF $0 \leq Q \leq Q_a$
Q_b APPLIES IF $Q_a < Q \leq Q_c$
Q_c APPLIES IF $Q_b < Q \leq Q_d$
Q_d APPLIES IF $Q_d < Q$
X_a APPLIES IF $0 \leq X \leq X_a$
X_b APPLIES IF $X_a < X$
ETC.

WHERE, Q^i = VOLUME (VEHICLES PER HOUR)
X^i, Y^i = PERCENTAGE (LEFT AND RIGHT TURNING MOVEMENTS RESPECTIVELY FROM THE i^{th} APPROACH LANE.

Figure H-2. Traffic parameters per approach lane.

volume counts and turning movements at the intersection, the optimal signal setting choice may be selected from the simulation results. However, the solution would be applicable only under those same traffic conditions. Another way to apply the solutions would be to have the control computer store them on magnetic tape or disk and select the one which corresponds to the traffic conditions, as currently sensed (Fig. H-3). Although this method is only of academic interest (since better results would be obtained from a traffic-actuated controller), it is practical in the sense that it would take only a small fraction of a second to obtain the stored solutions and implement them through commands sent over a communications network.

SEMI-TRAFFIC-ACTUATED CONTROL

The semi-traffic-actuated form of control is usually applied to an intersection of a heavy-volume (major) street with a relatively low-volume (minor) street. The signal is normally green on the major street and will respond to actuations on the minor street after a predetermined minimum green interval has elapsed. Hence, detectors placed only on the minor street may serve not only to change the signal from its normal state, but also to determine the length of the minor street's green phase within the limits of the predetermined minimum and maximum. Because there is no information concerning major-street traffic, the effective use of semi-traffic-actuated control is limited. The limitations for the case in which this special purpose hardware is installed would be determined at the time when the volume was heaviest on the major street. In the case of applying a computer program to implement the logic of the aforementioned hardware, the major-minor street volume conditions, obtained historically, would determine those portions of the day during which the control would be effective. In the control of traffic by the computer, many forms of control may be assigned to various combinations of signals (subnetworks) as a function of changes in the dynamic traffic pattern. Furthermore changes in the predetermined minimum, maximum, and extension intervals may be introduced as a function of current speed, roadway conditions, weather conditions, etc. This concept of applying current traffic control techniques by programming a digital computer to vary the mode of operation of any intersection and its interrelation with the surrounding signalized intersections presently is being applied in cities like Toronto, Canada, and San Jose, Calif.

Glossary

The following definitions are used in this section as applying particularly to conditions of semi-traffic-actuated control:

Major-street minimum green interval—the minimum amount of time during which the green signal indication is displayed to the major street before the right-of-way can be transferred to the minor street. If actuation occurs after this interval has expired, the right-of-way will be transferred immediately.

Major-street amber interval—the time allowed for the major-street traffic to clear the intersection before the movement of minor-street traffic.

Minor-street initial portion of green interval—the time allowed for queued vehicles to get into motion.

Minor-street extendible portion of the green interval—the amount of extension following each detector actuation after the initial portion.

Minor-street extension maximum—the limit set on the extendible portion of the green interval.

Minor-street amber interval—the time allowed for the minor-street traffic to clear the intersection before movement of the major-street traffic.

Operation

With no minor-street actuations the signal will remain green on the major street. When an actuation is detected the major-street amber interval will be displayed before the right-of-way is transferred to the minor street (after the expiration of the major-street minimum green interval).

Minimum minor-street green will be assumed to consist

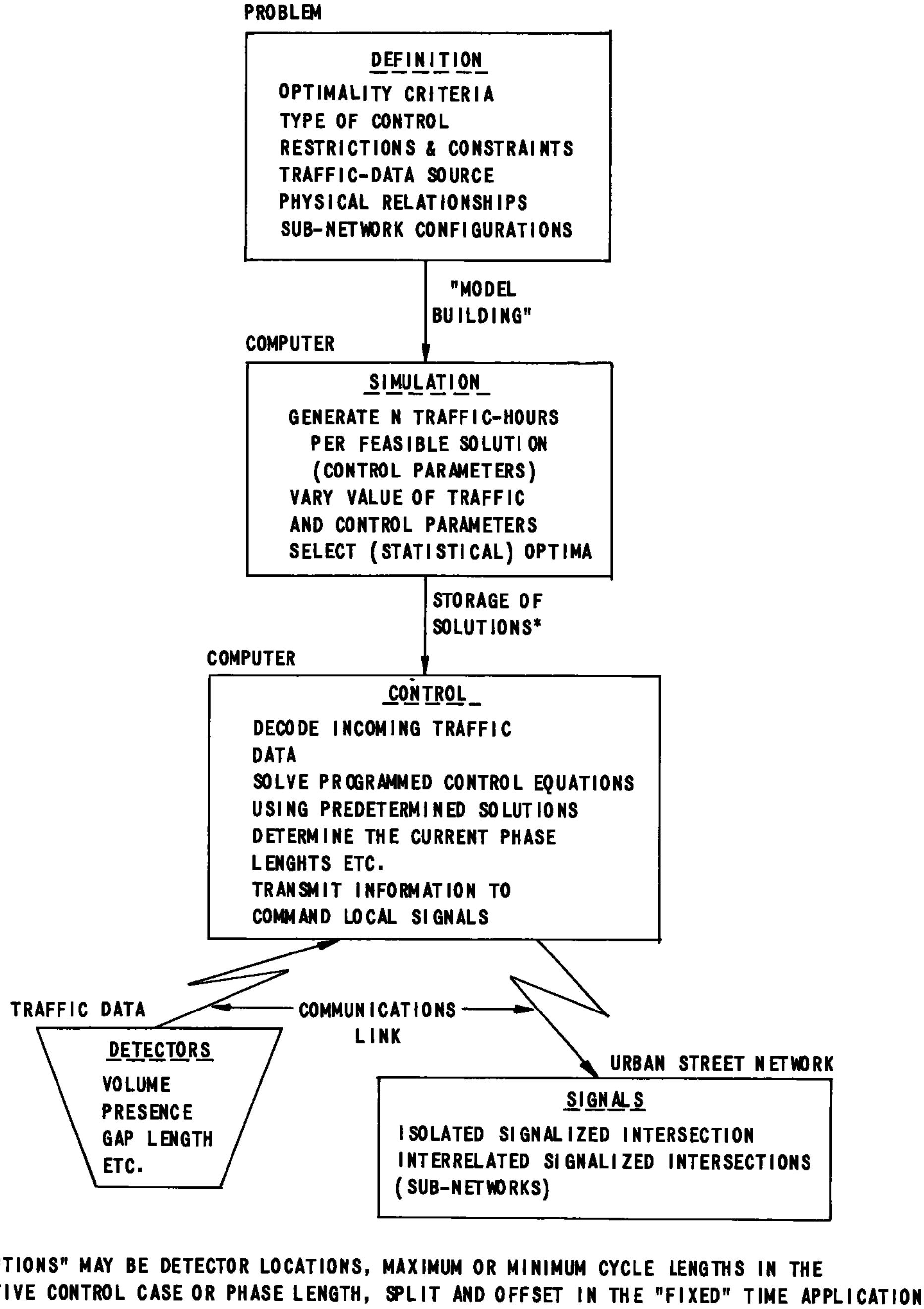


Figure H-3. Role of simulation in real-time control of traffic signals.

of an initial and an extendible portion of the green interval. Only actuations occurring during the extendible portion will affect the timing. Each succeeding actuation cancels the unexpired time of the previous extendible portion and adds another full extendible portion. This sequence of events (Fig. H-4) will continue until either an extendible period has elapsed without further actuation (decrease in the demand) or the minor-street maximum extension is exceeded. In either case the minor-street amber interval will be initiated before the right-of-way is transferred to the major street. In the latter case, the right-of-way will be returned to the minor street following the expiration of major-street minimum green interval. Also, the amber interval for minor-street clearance is extended in some available systems. The semi-vehicle-actuated intersection may be interconnected into a progressive system such that the minor-street green is only permitted in the system cycle when it will not disturb the predetermined artery progression.

Nomenclature

The following symbols are used in this section as applying particularly to the logic and programming for the conditions of semi-traffic-actuated control:

A = the major street;

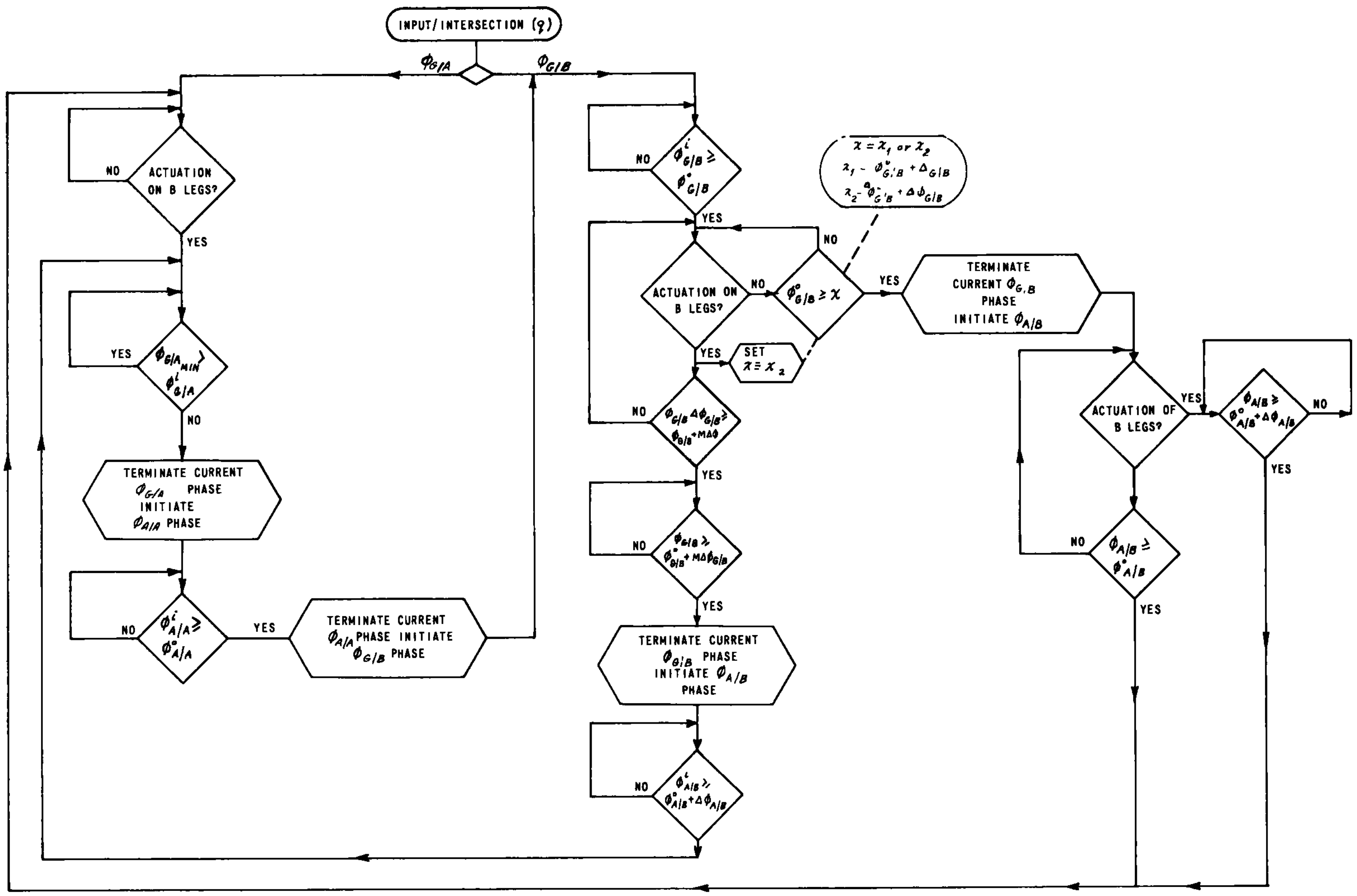


Figure H-4. Two-phase semi-vehicle-actuated controller.

- B = the minor street;
- $_q\phi^t_{G|A}$ = elapsed green time on major street, A, for intersection q at any instant;
- $_q\phi^t_{G|B}$ = elapsed green time on minor street, B, for intersection q;
- $_q\phi_{G|A\ min}$ = major-street minimum green interval (intersection q);
- $_q\phi_{A|A}$ = major-street clearance intreval (intersection q);
- $_q\phi^0_{G|B}$ = minor-street initial portion of green;
- $_q\Delta\phi_{G|B}$ = minor-street extendible portion of the green interval;
- $\mathrm{Max}_q(\Delta\phi_{G|B})$ = minor-street extension maximum or $M\Delta\phi_G|_B$;
- $_q\phi^t_{A|B}$ = minor-street clearance (elapsed time);
- $_q\phi^t$ = in general, the elapsed time of a particular phase, in seconds, referenced from time of initiation;
- $_q\phi^0_{A|B}$ = minimum clearance on minor street;
- $_q\Delta\phi_{A|B}$ = extension of minor-street clearance;
- $_qX_1$ = a substitution factor = $\phi^0_{G|B} + \Delta\phi_{G|B}$;
- $_qX_2$ = a substitution factor = $^a\phi^t_{G|B} + \Delta\phi_{G|B}$; and
- $^a{}_q\phi^t_{G|B}$ = time, referenced from the initiation of the phase, of the most recent actuation on B.

FULL-TRAFFIC-ACTUATED CONTROL

The full-traffic-actuated control is selected for intersections where traffic demands on all intersecting legs are considered necessary information in order to provide for the efficient operation of the signal. Detectors located on all the approaches to the intersection serve not only to assign the right-of-way but also to determine what the phase length should be. This phase length will fluctuate between the predetermined minimum and maximum values, depending on the length of time between actuations on the street having the right-of-way and the demand on the cross streets. Hence, the right-of-way is switched from street to street, depending on the frequency of gaps which exceed the predetermined time lengths for each street. As volumes increase gaps occur less frequently and the traffic-actuated signal operation approaches that of pretimed signals with the phase lengths equal to the predetermined maximums.

The added flexibility of allowing for the variability of the predetermined gap interval during a particular phase makes this form of control—namely, the traffic-density full-traffic-actuated control—much more sensitive to fluctuating traffic demands. For example, the decrease of the gap interval may be made proportional to the size of the queues waiting on the red signal, the increasing gap size as measured during the green phase, etc.

Glossary

The following definitions are used in this section as applying particularly to conditions of full-traffic-actuated control:

Initial portion of green interval—the time allowed for the queued vehicles to get into motion.

Extendible portion of the green interval—the time extension which follows each detector actuation after the initial portion has elapsed.

Extension maximum—limits the amount of time for the extendible portions of the green interval, after an actuation is detected on the opposite phase.

Amber interval—the time allowed to clear the intersection before the right-of-way is transferred to the other street.

Operation

When the right-of-way is first transferred, the minimum green time which must be displayed consists of an initial interval and one extendible interval. Detector actuations occurring during the initial interval have no effect on the timing. Actuations occurring during the extendible interval will cancel the unexpired time of the previous extendible interval and add another full extendible interval. If no actuations are detected on the cross street, this sequence of events continues indefinitely (Fig. H-5). When a detector actuation occurs on the cross street, the right-of-way will be transferred at one of the following times:

1. Immediately if the extendible interval has expired.
2. At the expiration of the current extendible interval if no other actuations are detected before it expires.
3. At the expiration of the maximum green interval if successive actuations are received on the street which now has the right-of-way and they are separated by less than the extendible interval.

It should be noted that the maximum green interval is referenced (zeroed) from the time of the first actuation on the cross street. If the current phase is terminated due to the expiration of the maximum green time, the subsequent amber interval will be extended and will be returned to the initial state after timing maximum green (as if the last actuation had occurred during the initial period).

Nomenclature

The following symbols are used in this section as applying particularly to the logic and programming for the conditions of full-traffic-actuated control:

- $\phi^0_{G|A}$, $\phi^0_{G|B}$ = initial portion of the green interval for legs A and B, respectively;
- $\Delta\phi_{G|A}$, $\Delta\phi_{G|B}$ = extendible portion of the green interval for legs A and B, respectively;
- Max $(\Delta\phi_{G|A})$, Max $(\Delta\phi_{G|B})$ or $M(\Delta\phi_{G|A})$, $M'(\Delta\phi_{G|B})$ = extension maximum for the extendible portion of the $\phi_G|_A$ and $\phi_G|_B$ phase, respectively; considered only after an actuation on the opposite legs;
- $\phi^0_{A|A}$, $\Delta\phi_{A|B}$ = minimum amber interval for legs A and B, respectively;
- $\Delta\phi_{A|A}$, $\Delta\phi_{A|B}$ = an extension of the minimum clearance interval (extended only if maximum timer has operated or if detector actu-

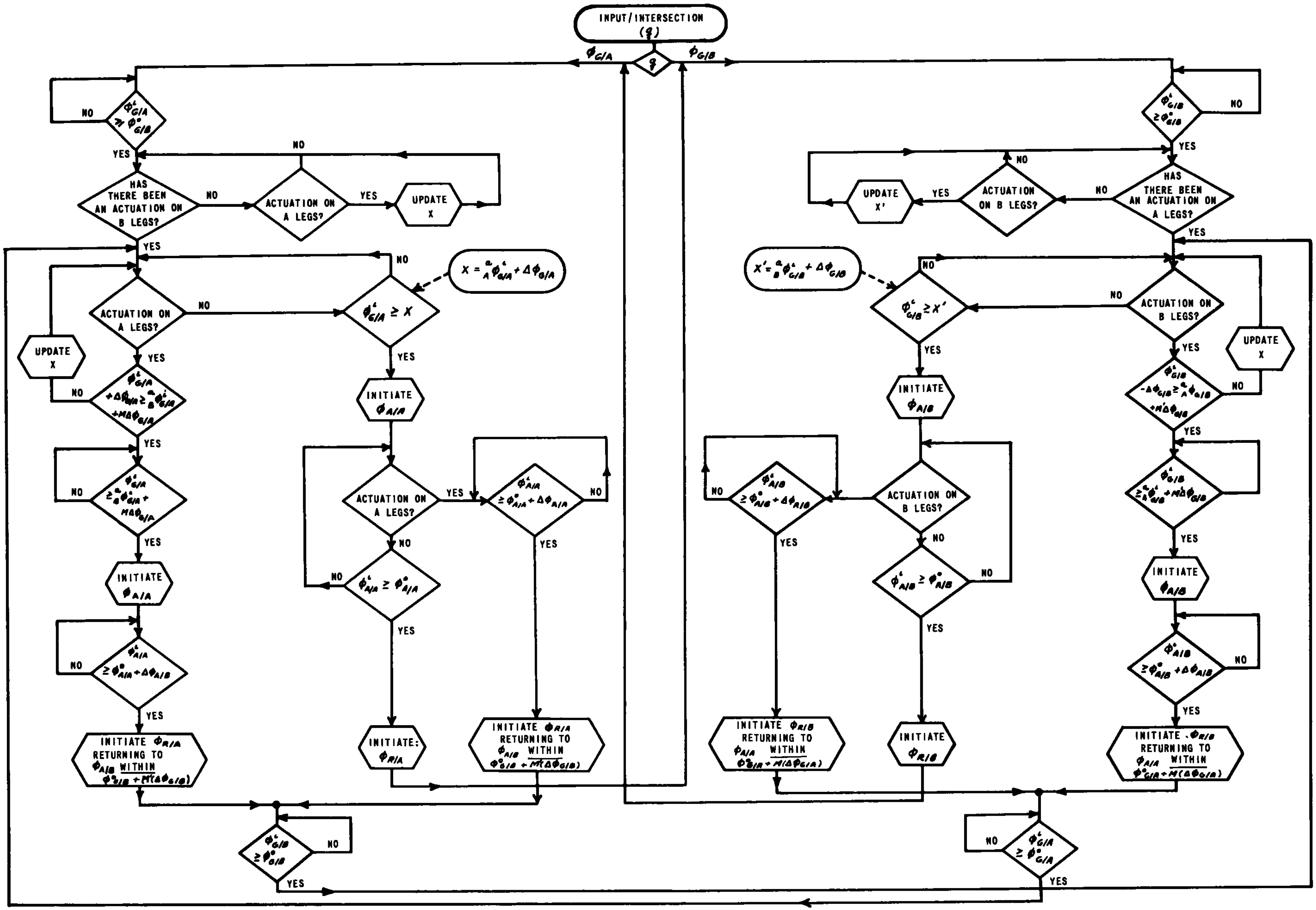


Figure H-5. Two-phase full-vehicle-actuated controller.

ation is received during clearance interval); and

$^{a}{}_{A}\phi^{i}{}_{G}, {}^{a}{}_{B}\phi^{i}{}_{G}$ = time of actuation (into green phase) on legs A and B, respectively.

COORDINATED CONTROL SYSTEMS

The object of coordination is to establish a relationship between the start of the green phase at two or more intersections along a major street or in a network of streets.

There are several forms currently available for linking the intersections on a main traffic route. The simultaneous, or synchronized, system and the alternate, or limited progressive, system are two of the most elementary forms of coordination. In the former case all signals along the controlled street display the same phase to the same traffic stream at the same time. In the latter case adjacent signals present opposite phases alternately along a particular street. Variations on the systems may be made by applying the aforementioned progressive systems to adjacent groups of signals.

However, the progressive system, as it is commonly called, provides for the cycle time for each intersection in the system to be fixed, although the green phases are displaced with respect to each other according to the desired progression speed. This system may be used to give preferential movement in favor of inbound flow at the expense of the minor flow traveling in the outbound stream, or vice versa, as conditions warrant. In general, the series of signals to be coordinated may be of the semi-traffic-actuated type, in which case a supervisory background cycle is imposed on the signals. This serves to assure that each of the semi-actuated controlled intersections provides a minimum green phase in a time relation (offset) best suited for the progression speed in a particular direction. In the case where there is no demand detected on the minor street, those intersections are permitted to operate as isolated vehicle-actuated signals contingent upon the condition that there is no interference with the prescribed plan of progression. The following is a description of a special purpose system similar to the one now in operation in Buffalo, N. Y.

Operation

Control decisions are based on lane occupancy (*18, 22*), a factor obtained from the common presence detector by calculating the ratio of the total accumulated vehicular passage time measured during a predetermined time interval to the total elapsed time of the interval. It is believed (*22*) that such a quantity results in a more significant measure of congestion than does density. The system is designed to control the cycle length, offset, and split for a number of signals along a major arterial. Four detectors are proposed to obtain the necessary control measurements, including two detectors sampling inbound traffic and two detectors sampling outbound traffic.

In general, the common cycle length will be selected by considering the level of the calculated lane occupancy, and the offset will be chosen as a function of the relative values of inbound and outbound lane occupancies. Such a relative indicator of the predominant flow, or more precisely the congestion, may be obtained by taking the ratio of the inbound or outbound lane occupancy to their sum.

Hence, the indicator may be expressed as $I/(I+O)$, where I and O represent the inbound and outbound lane occupancies, respectively. For example, if $I > O$ then the comparative term becomes greater than 0.5; similarly, if $I = O$ then the term is precisely equal to 0.5. From predetermined classifications the system will operate on either an average or a preferential offset mode. That is, for some value of $I/(I+O) > 0.5$ the system will select an inbound offset, for some value of $I/(I+O) < 0.5$ the system will select an outbound offset, and for the values between an average offset will be implemented. The only exception is when the values of lane occupancy for both inbound and outbound traffic fall below a predetermined minimum. This restriction does not permit the system to fluctuate between inbound and outbound offsets for small traffic variations, and therefore adds to performance stability.

Having selected one of three offsets (inbound, outbound, or average), the computer is prepared to select the common cycle length (Fig. H-6). For the case of preferential offset modes the computer calculates the total duration of the cycle to be implemented by computing the level of lane occupancy determined by averaging the two sampling detector outputs from either the inbound or outbound lanes. In the case of average (nonpreferential) offset the computer accepts information from each of the four detectors and calculates an average value using the inbound and outbound occupancy measurements. Based on the value obtained as a result of the aforementioned lane occupancy averages, the computer will decide to which of six predetermined levels of lane occupancy the calculated average belongs. This then determines the cycle length, as previously assigned to each level.

Lane occupancy as previously defined is equivalent to the measurement of the percentage of time during which a vehicle is sensed by the detector. Because these measurements must themselves be averaged over a particular interval of time, say x minutes, this average may be denoted as a moving average which will determine the percentage of the last x minutes during which the associated detectors were occupied. The resultant values of lane occupancy may be made more sensitive (quicker response) to the state of increasing congestion than to that of decreasing congestion by permitting different adjustments for the sampling interval. Further constraints that provide for over-all insensitivity to instantaneous fluctuations in traffic (greater operational stability) are to set values for the maximum frequency of changes in offset and require the system to pass through the average mode when transferring between the inbound and outbound modes.

Nomenclature

The following symbols are used in this section as applying particularly to the logic and programming for the conditions of coordinated (progressive flow) systems:

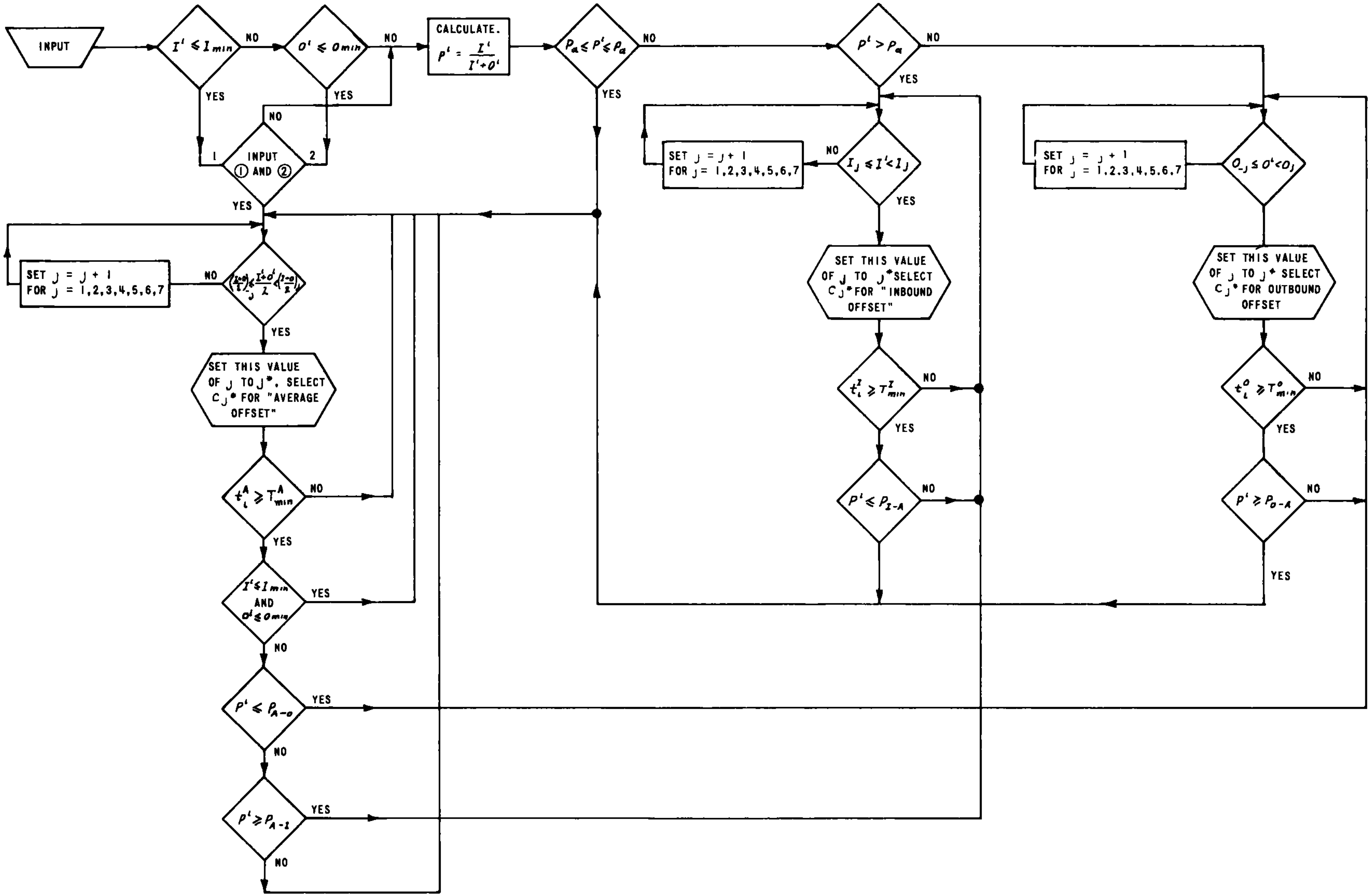


Figure H-6. Progressive mode; selection of offset and cycle length.

I^i = average inbound lane occupancy;
O^i = average outbound lane occupancy;
$\left(\frac{I+O}{2}\right)_{j \text{ or } -j}$,
$I_{j \text{ or } -j}$, $O_{j \text{ or } -j}$ = predetermined ranges from which to select cycle lengths, where j is one of 7 levels of occupancy, yielding 6 cycle lenths;
C_k = cycle length, $k = 1, \ldots, 6$;
$C_k{}^*$ = chosen cycle length for a particular progressive mode;
$t_i{}^A, t_i{}^I, t_i{}^o$ = elapsed time of the current average, inbound, or outbound progressive modes, respectively;
$T_{\min}{}^A, T_{\min}{}^I$,
$T_{\min}{}^o$ = minimum time for the average, inbound, and outbound progressive modes;
$I_{\min}, O_{\min}$ = values of the inbound and outbound lane occupancies, respectively, such that the average offset mode is automatically selected if I^i and O^i fall below;
$P^i = I^i/(I^i + O^i)$;
P_{-a}, P_a = threshold values that initially determine the offset mode based on the calculated P^i; and
P_{A-O}, P_{A-I},
P_{I-A}, P_{O-A} = threshold values which determine (after the minimum elapsed time) the value of P^i necessary before the offset mode is changed from average to outbound, average to inbound, inbound to average, and outbound to average, respectively.

CONCLUDING REMARKS

The main purpose of the preceding discussion has been to comment academically on the application of traffic simulation to control, and to present a few of the most commonly implemented traffic control strategies. Although these control doctrines are universally* employed through the use of special purpose analog hardware (*22, 23, 24*), their adaptation to the general purpose digital computer begins with the flow-charting of the control logic. These charts have been presented in Figures H-4, H-5, and H-6 for the particular modes described.

The same control modes could be programmed for the computer and, through an analysis of the street and traffic patterns, assigned to various subnetwork configurations for the city being considered. This would serve as a model for costing the control system designed to emulate conventional control strategies.

To have been consistent with the study performed for a digital-computer-controlled system using an advanced adaptive control logic as given elsewhere in this report, these costs should have been obtained. This course has not been pursued, the principal reason being that the capabilities of the digital computer are not deemed as being adequately exploited, if it is used only to implement the more conventional control modes.

* Exceptions may be found in the cities of Toronto, Canada, and San Jose, Calif.

APPENDIX I

ANNOTATED BIBLIOGRAPHY

As part of the synthesis of a digital-computer-controlled signal network a number of documents dealing directly with the subject of traffic control and also with other supporting subjects (statistics, computers, electronics, vehicular dynamics, etc.) have been read with varying interest and completeness. It is the purpose of this bibliography to present a large portion of these documents and comment briefly on their contents. A word of caution is in order, inasmuch as these comments are not intended to be either academically objective, in the form of a book review, or academically subjective, in the form of a critique, but rather were written with a bias (i.e., a particular goal) as the documents were read. This goal represents the objective of the project (i.e., to synthesize a real-time digital-computer-controlled signal system for a city such as White Plains, N.Y.) If, indeed, the bibliography and comments are of interest to others, this is a by-product and not the intention of the author.

ACTON, F. S., *Analysis of Straight-Line Data.* Wiley (1959).

This book was written for the "statistically self-educating physical scientist and engineer" who deals with data containing one or more lines entrapped within his experimental variability. Emphasis is placed on the analysis of variance as a mathematical tool. Deliberately omitted are some of the classical approaches that encourage an unthinking mechanical solution to the problem of straight-line fitting.

BECKMANN, H., MCGUIRE, C. B., and WINSTEN, C. B., *Studies in the Economics of Transportation.* Yale Univ. Press (1955).

This book is addressed to analysts in various professions, including economists, traffic and railroad engineers, management scientists, operations researchers, and mathematicians. Studies relating to highway traffic and railroad transportation are offered.

Of particular interest to the project is Chapter 1, entitled "Road and Intersection Capacity," which applies queuing theory to such traffic situations as the flow of cars through an intersection and the passing of slower cars by faster ones using gaps which occur in the opposing traffic stream.

BERKELEY, E. C., and WAINWRIGHT, L., *Computers, Their Operation and Applications.* Reinhold (1956).

The purpose of this book is to present basic information about computers, particularly automatic computers. The authors realize that the development in the field is so rapid that some of the information in this book will shortly become out of date, but many of the concepts and discussion of applications are still current. Of particular interest was Section II, entitled "Automatic Digital Computing Machines."

BONE, A. J., MARTIN, B. V, and HARVEY, T. N., "The Selection of a Cycle Length for Fixed-Time Traffic Signals." Joint Highway Research Project, MIT Dept. of Civil Engineering and Mass. Dept. of Pub. Works (Apr. 1964).

This paper describes a general method for determining the optimum cycle length for fixed-time signalized intersections, using a combination of three criteria to evaluate various cycle lengths. The criteria are the delay per vehicle, the expected queue length, and the probability of entering the intersection during the first green phase. Two examples are used to illustrate the application of the method and the use of the formulas. A computer program for calculating the aforementioned criteria is described. The techniques utilized were primarily due to Webster's work on signal timing, including the formula for the delay per vehicle. It is noted by the authors that the engineer must make the final decision as to the relative importance attached to each criterion at a particular location.

CASCIATO, L., and CASS, S., "Pilot Study of the Automatic Control of Traffic Signals by a General Purpose Electronic Computer." *HRB Bull. 338,* pp. 28-39 (1962).

This paper describes the system concepts employed in a pilot study designed to control a network of traffic signals (9 to 16 signalized intersections) by a general purpose digital computer in Metropolitan Toronto. The purpose of the study was to demonstrate that the computer could be connected to an existing traffic signal network to provide a greater degree of service and flexibility than was heretofore attainable with the conventional (special purpose) control methods. Also, it was desired to learn from the study how the system could be used to improve traffic flow, measure such improvement, and provide data for post-analysis requirements. Vehicle detectors which were installed on most approaches and existing traffic signals were connected to the central computer by telephone cables. The paper briefly discusses the master control program stored in the computer which may assign various plans which in turn define a mode of operation for the particular signals. However, no details are given as to the programs used, or the relative success of individual control philosophies that were simulated or newly developed for the system. Also included is a discussion of platoon structure, delays and congestion within the limitations of this particular system's instrumentation. Results are given in terms of the average delay per vehicle and the defined "congestion." It should be noted that this performance was compared with an existing fixed-time system with no indication as to the nature of the particular signal settings or network considerations.

COX, D. R, "Prediction by Exponentially Weighted Moving Averages and Related Methods." *Jour. Royal Stat. Soc.,* Series B, Vol. 23, No. 2, pp. 414-22 (1961).

This paper is a highly technical (mathematical) development which assumes previous background in statistics. The paper calculates the mean square error of prediction for an e.w.m.a. when the series is a Markov series. The optimum choice of the damping constant is given, although not critical. There is a value of the Markov correlation below which it is impossible to predict the local variations of the series with an e.w.m.a. A modified e.w m.a. having a mean square error approaching that for the best linear predictor (Wiener) is formulated. This is of value if the Markov correlation parameter is negative and possibly when the correlation parameter is near the limit ρ_0.

FRENCH, R. A., "Coordinated Traffic Signalling System—Sydney, Australia." *Traffic Quarterly,* pp. 76-88, The Eno Foundation for Highway Traffic Control. (Jan. 1965).

This paper states that a control system responding to vehicle detectors (as presently available) is not adequate for the central business district. Particularly, it is argued that closed-circuit television, coupled with trained observers, is the most convenient and economical method for systems having about 100 control points or local intersections. A system is described in detail, consisting of twelve cameras strategically mounted on high buildings in the central business district of Sydney, Australia. The basis of the traffic control system is the coordinated operation of traffic signals at local intersections to minimize delay, consistent with safety requirements. The conclusions indicate that the installed system is operational and gives the results of the (desired) reduced travel times. However, it is recognized that control depends on the quality of traffic condition assessments made by the operator, and his choice of control programs. There is a limit to the amount of information that the operator can absorb and process at any one time, and this could limit the efficiency of control achieved. The paper further states that vehicle detectors have been placed experimentally in two sections of a two-way street.

GARWOOD, F., "The Sampling and Use of Traffic Flow Statistics." *Applied Statistics,* Vol. 11, No. 1 (1962).

This article discusses two main sampling problems which arise in traffic flow statistics. The first is that of the accuracy with which trends from year to year in the total amount of travel on the road can be estimated from continuous counts at a sample of points on the road system. The second is concerned with sampling in time at particular points and the effect of geographical factors on sampling efficiency. One application of this concerns the concept of the economic capacity of the road, the formula for which involves the coefficient of variation of the hourly flows. Values of the latter are given for some typical roads.

GAZIS, D. C., "Optimum Control of a System of Oversaturated Intersections." *Jour. Oper. Res. Soc.*, Vol. 12, No. 6 (Nov.-Dec. 1964).

The problem of optimizing the control of two oversaturated intersections is solved by semigraphical methods. An analytical formulation of the method using Pontryagin's control theory is also given. It is concluded, however, that it is not clear that the control theory presented could be useful for a system of more than a few intersections. Particularly, some experimentation is still needed to establish the feasibility of the proposed method in the case of the single intersection.

GAZIS, D. C., and POTTS, R. B., "The Oversaturated Intersection." International Business Machines Corp.

The situation examined in this paper is that of two competing demands which exceed the capacity of the intersections. Therefore, for the given condition queues are formed and maintained during this period of flow (oversaturation). A method (see Gazis, p. 8) is presented for optimizing the control of an oversaturated intersection by a traffic signal to produce a reduction in delays.

GERLOUGH, D. L., "Some Problems in Intersection Traffic Control." *Theory of Traffic Flow*, Robert Herman (Editor), Elsevier (1961).

This paper examines some of the problems underlying the control of traffic at intersections, the existing intersection control systems, and the areas in which further theoretical study and equipment improvement are required, and presents a progress report on the attack on one of the problems. The discussion is in the terminology and methods of analysis used in connection with automatic control systems. It is pointed out that existing traffic control systems at intersections are largely open-loop in nature (i.e., minimum feedback) and that to achieve the greatest efficiency techniques must be developed for measuring and controlling (an agreed-on figure of merit) on a closed-loop basis. Further interim solutions to platoon problems in the form of mathematical representations of platoons have been proposed.

GERLOUGH, D. L., and CAPELLE, D. G. (Editors), "An Introduction to Traffic Flow Theory." *HRB Spec. Report 79* (1964).

This publication is an attempt to consolidate recent contributions to the theory of traffic flow and present the material in a related manner with a consistent notation. It is noted that many of the theoretical descriptions presented in this paper have not been completely validated and verification or refinement is necessary before the theories can become useful analytic tools. Of particular interest to this project is Chapter Three, entitled "Queuing Theory Approaches."

GERLOUGH, D. L., and SCHUHL, A., *Poisson and Traffic.* The Eno Foundation for Highway Traffic Control (1955).

This paper contains two mathematical discussions of traffic analyses; namely, "Use of Poisson Distribution in Highway Traffic" and "The Probability Theory Applied to Distribution of Vehicles on Two-Lane Highways." In the first portion, the basic distribution is considered and then applied to a number of typical traffic problems. The results of the second lead to a closer agreement between theory and observations than has been obtained heretofore (e.g., the probability of a spacing larger than x, probability that an interval taken at random contains no vehicles, but is bounded by a vehicle on one side, etc.)

GERLOUGH, D. L., and WAGNER, F. A., "Improved Criteria for Designing and Timing Traffic Signal Systems." Final Report on NCHRP Project 3-5, Planning Research Corp. (Mar. 1964).

Before the formulation of a model to study traffic at individual intersections under laboratory conditions was undertaken, all known past and present research having a bearing on the subject was reviewed. As a result, a comprehensive bibliography of 282 entries was prepared and abstracts of the most significant papers (28 entries) are presented along with the bibliography in an appendix to this report. The model developed for the computer considers in detail the behavior of each vehicle in the system (i.e., microscopic). Numerous graphs of vehicular behavior are presented, together with flow charts of the various computational subroutines contained in the model. Future plans include the investigation of various measures of effectiveness for intersection control, study of the effectiveness of existing control techniques, and description and testing of new traffic signal control techniques.

GILBERT, K., "Evaluation of the Six-Minute Sample Count Procedure." *Traffic Eng.*, pp. 17-20 (Nov. 1962).

This paper describes the 6-min sample count test procedure and presents the results. The test had the goals of (a) determining the accuracy of the 6-min sample count procedure over a wide range of volumes, and (b) preparing data in useful form as a tool for selecting applications and evaluating results of the 6-min count procedure. This count procedure is a method of estimating existing traffic volumes by taking a 10 percent sample manual count each hour over the desired counting period.

GRACE, M. J., and POTTS, R. B., "A Theory of the Diffusion of Traffic Platoons." *Jour. Oper. Res. Soc.*, Vol. 12, No. 2, pp. 255-275 (Mar.-Apr. 1964).

This paper presents a theoretical study of the diffusion of a traffic platoon as it travels down a road. It is assumed in the mathematical model that the speeds of the vehicles in the platoon are distributed normally. The parameters of this normal distribution are expressed in terms of a measure of the spreading of the platoon (i.e., diffusion constant). The problem of coordinating two successive traffic lights is solved for a number of assumed initial conditions. In the

application of the model to the problem of the coordination of two successive traffic lights, the initial density distribution function was chosen as a rectangular or trapezoidal pulse on the basis of experimental results. Some of the important limitations of the model as applied to practice are (a) the theory is limited to traffic of medium volume; (b) in the case of three or more lights it is not clear whether the model can be used from signal to signal, or from first to last signal; (c) right or left turning has not been considered; and (d) no effects of amber phase have been considered.

GRAHAM, E. F., and CHENU, D. C., "A Study of Unrestricted Platoon Movements of Traffic." *Traffic Eng*, Vol. 32, No. 7, pp. 11-13 (Apr. 1962).

This study was performed at a site on US 40 between San Francisco and Sacramento, by the California Division of Highways, in cooperation with the Bureau of Public Roads. The objective was to determine the amount of dispersion of platoons released from a traffic signal with minimal downstream interference. The study concluded that the traffic remained in fairly well-defined platoons for distances of at least 1 mile beyond the rural, isolated, signalized intersection.

GREENBERG, H., and DAOU, A., "The Control of Traffic Flow to Increase the Flow." *Jour. Oper. Res. Soc.*, Vol. 8, No. 4, pp. 524-532 (1960).

An operational study of tunnel traffic is presented. Data of flow and density at the bottleneck are fitted to a fluid model description of the flow. The control of the input traffic by platooning is shown to eliminate shock waves and permit higher flows. Experimental results are shown to justify the theory and methods presented.

GREENSHIELDS, B. D., SCHAPIRO, D., and ERICKSEN, E. L., "Traffic Performance at Urban Street Intersections." *Tech. Rep. No. 1*, Yale Bur. of Highway Traffic (1947).

This report studies the traffic movements of individual vehicles at both signalized and unsignalized urban intersections. The data were acquired by time-lapse photography. Hence, a complete record of events that occurred within the field of view was recorded and in some instances correlated to the particular vehicular movements (e.g., comparison of reaction time between successive vehicles, showing the effect of the intensity of pedestrian traffic.) Further, since it was possible to distinguish between vehicle classifications, a distinction in acceleration and deceleration profiles was made for trucks, buses, and passenger cars. A major effort was made to record starting reaction times to signal changes and between successive vehicles at a number of intersections. Although the emphasis of the study is on the transients associated with the intersection (e.g., acceleration, deceleration, and reactions), measurements were made during the steady-state condition (i.e., car following). Graphs of time and distance clearances versus speed are presented. It should be noted that this study was initiated in 1944 and care should be exercised in applying certain of the empirical, numerical results.

GREENSHIELDS, B. D., and WEIDA, F. M., *Statistics with Applications to Highway Traffic Analyses.* The Eno Foundation for Highway Traffic Control (1952).

This book presents a methodical discussion of some statistical theories and their application in the analysis of traffic data. It is considered by the authors as an introduction to the subject. The first four chapters explain the mathematics as a tool, and the final chapter shows its application.

GRIMSDALE, R. L., MATHERS, R. W., and SUMNER, F. H., "An Investigation of Computer-Controlled Traffic Signals by Simulation." *Proc. Inst. Civil Eng.*, Vol. 25, pp. 183-191, Paper 6645 (1963).

This paper describes some experiments in computer-controlled traffic signals contained in a particular road network. The network and the control philosophies were simulated by the Mercury digital computer at Manchester University. The object of the experiments was to investigate some methods of centralized computer-control of traffic signals and in particular the provision of signals indicating the best routes between points in the network. Results indicated that simple fixed-cycle control was the least efficient, whereas the most efficient control doctrine incorporated routing control combined with a form of restricted entry. Further discussions on practical difficulties point out that an extensive network of measuring devices and communication lines is required to place the traffic signals of the network under the control of a central computer. Also, the basic measuring devices (vehicle presence and vehicle flow counters), would have to be placed at selected points along the lane to determine the number of vehicles in each lane. The authors suggest that a number of checks could be used to correct some of the errors that would undoubtedly occur.

HAIGHT, F. A., *Mathematical Theories of Traffic Flow.* Academic Press (1963).

This book attempts to justify the theory of traffic flow as a part of applied mathematics. Particularly, the author attempts to bring out the fundamental relationship between traffic flow theory and the classical subjects of queuing theory, stochastic processes, and mathematical probability. Although the terminology of roads and vehicles is employed, many parts of the book have wider application.

HEWTON, J. T., "The Metropolitan Toronto Automatic Traffic Control System" Traffic Engineering Dept., Municipality of Metropolitan Toronto.

This report describes Metropolitan Toronto's automatic traffic control system as proposed by the Traffic Research Corporation (see Casciato and Cass). A fairly detailed description of the components (computer, peripheral equipment, detectors, signal modification unit, communications hardware, monitor and public display units) is included. Detailed discussions of the installation procedure and the practical considerations are given The breakdown of the costs is for a complete system comprising 1,000 signalized intersections and 2,000 vehicle detectors, although

Toronto presently has only 600 signals. A portion is devoted to explanations and diagrams of the computational procedures used for various modes of control.

"Highway Capacity Manual—1965." *HRB Spec. Report 87* (1965).

The combined efforts of many organizations and individuals applied in a number of locations for many years have resulted in a great mass of field observations. As a consequence of such data and subsequent analyses, this manual presents the capacities of rural highways with uninterrupted flow, the capacities of intersections at grade, weaving intersections, grade separations, and ramps, and the relation of hourly to annual average traffic volumes. This document is meant to be a practical guide by which the engineer can design a highway or revamp an old one "with the assurance that the resulting capacity will be as calculated." One chapter is devoted to the definitions of numerous terms which are necessary and fundamental to the particulars which follow.

HILLIER, J. A., "A Review of Developments in Area Traffic Control." *Proc. First Conf. of the Australian Road Research Board,* Vol. 1, pp. 416-429 (1962).

This paper describes some of the more important existing forms of coordinated area traffic control (automatic diversion schemes and large-scale linking of signals). Also included is a description of a number of proposals which have been made for future development and the research currently being conducted at the Road Research Laboratory.

HOEL, P. G., *Introduction to Mathematical Statistics,* 3rd Ed. Wiley (1962).

This book considers the theory and application of statistics simultaneously, although throughout the text the emphasis is on the theory. In terms of solving problems in statistics, this book emphasizes the selection of a mathematical model and the conclusions drawn from the model to solve the proposed problem. However, no details are given as to the tests for the reasonableness of the model.

HOLT, C. C., MODIGLIANI, F., MUTH, J. F., and SIMON, H. A., *Planning Production, Inventories and Work Force.* Prentice-Hall (1960).

This book undertakes to describe some decision-making mathematical techniques and how they may be applied to managerial decisions in the operation of a factory warehouse system. Although the methods reported were developed in the context of a factory supplying a warehouse system, they are applicable to the corresponding decision problems in the operation of military and other governmental organizations. Moreover, the basic mathematical tools can be adapted to use in other fields. Of particular interest is Chapter 14, entitled "Forecasting Sales of Individual Products by Exponentially Weighted Moving Averages," which includes details on the methods of forecasting (including modifications to take account of seasonal and trend effects), selection of weights and initial value, and tests of the forecasting methods in order to evaluate the predictive accuracy of the exponential system.

KELL, J. H., "Analyzing Vehicular Delay at Intersections Through Simulation." *HRB Bull. 356,* pp. 28-39 (Jan. 1962).

This paper describes the development of a model for the simulation of an intersection of two two-lane bi-directional streets, with stop signs controlling one street. The studies made to obtain adequate mathematical distributions describing traffic behavior are discussed. Also presented are the simulation results along with the formulated relationships between vehicular delay and approach volumes and turning movements.

KELL, J. H., "Intersection Delay Obtained by Simulating Traffic on a Computer." *Highway Research Record No. 15,* pp. 73-97 (1963).

The intersection model used for the simulation results presented in this paper is an orthogonal intersection of two, two-lane, two-way streets. The intersection operation has been described for both stop-sign and signal control. The results (total delay) are presented graphically for the stop-sign and fixed-time signal for various conditions (e.g., volumes, signal splits, etc.)

KELL, J. H., "Results of Computer Simulation Studies as Related to Traffic Signal Operation." Paper presented at 33rd Ann. Meeting, Inst. of Traffic Engineers, Toronto (Aug. 1963).

This paper describes the IBM 7090 computer program designed to simulate traffic at an intersection controlled by traffic signals. The form of the signal control may be pre-timed, semi-actuated, or full-actuated. The results obtained from the simulation of the pre-timed signals only, are presented in great detail in the form of graphs (e.g., total delay vs volume for various cycle lengths and splits, etc.) It should be noted that the emphasis of this paper is on the results of the simulation, with 25 pages devoted to graphical presentation.

LEISCH, J. E., "Design Capacity Charts for Signalized Street and Highway Intersections." *Pub. Roads,* Vol. 26, No. 6.

This paper presents a method of graphic analysis based on Part V of the *Highway Capacity Manual* (1950 edition), which describes a method of analysis for the calculation of capacity at signalized intersections under various conditions. The design capacity charts are presented with typical examples for various types of signal-controlled intersections on two-way streets, one-way streets, and expressways.

LITTLE, J. D. C., MARTIN, B. V., and MORGAN, J. T., "Synchronizing Traffic Signals for Maximal Bandwidth." Dept. of Civil Engineering, MIT (Mar. 1964).

This paper contains a description of an IBM 1620 computer program (a listing and operating instructions are included) developed for the solution of two particular traffic problems. The first is: Given an arbitrary number of signals along a street, a common cycle length, the green and red times for each signal, and specified vehicle speeds in each direction between adjacent signals, synchronize the signals to produce bandwidths that are equal in each direction and as large as possible. The second is to adjust the synchronization to increase one bandwidth to some specified feasible value and maintain the others as large as possible. The definition of bandwidth is given as "that portion of a signal cycle for which it is possible for a vehicle starting at one end of a street and traveling at preassigned speeds to go to the other end without stopping for a red light."

MATSON, T. M., SMITH, W. S., and HURD, F. W., *Traffic Engineering.* McGraw-Hill (1955).

This book attempts to explain highway traffic phenomena by simple and elementary analyses. The book is divided into five sections entitled, "Characteristics," "Regulations," "Control Devices and Aids," "Design," and "Administration and Planning." Of particular interest is the discussion of vehicle characteristics and traffic signals.

MILLER, A. J., "A Computer Control System for Traffic Networks." *Second Internat. Symposium on Theory of Road Traffic Flow,* Univ. of Birmingham (England) (June 1963).

This paper describes a method of controlling traffic signals employing the minimum-delay criterion. The implementation requires a digital computer fed information from strategically located detectors. With the control system described, the computer interrogates all the intersections in the network within a predetermined time interval to make the decision to have the current signal state unchanged for some integral multiple of the time interval, or change the signals immediately. The decisions are based on measurements and/or estimates of queue lengths, departure and arrival rates, phase duration, etc. Detectors are proposed at stop-line and midblock locations for each approach. The emphasis is on the individual intersection, although there is a discussion of network considerations. Also included are comments on exponentially weighted moving averages (as a prediction technique), fixed-time synchronization, quantity control (limiting queue lengths), and route control.

MILLER, A. J., "Computers in the Analysis and Control of Traffic." Dept. of Transportation, Univ. of Birmingham (England). Paper prepared for a conference on Computers in Civil Engineering.

This paper attempts to justify the use of computers in the control of traffic. The justification is sought by finding control strategies which are superior to those available with existing controllers. That is to say, strategies that could appreciably reduce delays over what could optimally be expected from conventional techniques and equipment. The paper briefly discusses various forms of computer control but only considers timing control in detail. The isolated intersection is developed, then the network case. Also pointed out are the limitations of such machine control, and particularly its dependence on sensory devices (i.e., detectors and human observations).

MILLER, A. J., "Settings for Fixed-Cycle Traffic Signals." *Oper. Res. Quart.* (Dec. 1963).

This paper presents the full mathematical details of the derivation of a formula to represent the average delay to vehicles at junctions controlled by fixed-cycle traffic signals. The delays predicted are close to those observed in practice. The delay formula is used to derive optimum signal settings.

NATIONAL JOINT COMMITTEE ON UNIFORM TRAFFIC CONTROL DEVICES, *Manual on Uniform Traffic Control Devices for Streets and Highways.* U. S. Dept. of Commerce (June 1961).

This manual sets forth the basic principles that govern the design and use of traffic control devices. These principles appear throughout the text in discussions of the devices to which they apply. The standards apply to any and all streets. Where a device is intended for limited application only, or for a specific system, the text specifies the restrictions on its use. Although the manual describes the application of the various devices, it is not intended as a substitute for engineering judgment. Both engineering judgment and imaginative application are essential to true uniformity.

OLIVER, R. M., "Travel Times Through Shock Waves." *Oper. Res. Quart.,* Vol. 15, No. 2 (June 1964).

This paper formulates the principle of conservation of matter with an integral equation which expresses travel times in the traffic stream. A particular situation is presented whereby the flow rates into the bottleneck (which are time dependent) temporarily exceed the capacity of the bottleneck. Expressions are presented for queue sizes, location and velocity of shock waves, and delays to vehicles in the traffic stream.

"San Jose Traffic Control Study, Initial Report." Data Processing Div., IBM, Kingston, N. Y.

This report contains a description of a particular digital-computer control system, including details of the management and hardware necessary for its implementation. Computational procedures of various subroutines are presented, with a functional description of the system's operation and specifications. The study will begin with known methods of control in Phase I and work toward implementing improved methods of arterial control; plans for Phase II include the more difficult techniques for grid control.

SCHENLER, W. W., and MICHAEL, H. L., "A Sufficiency Rating Method for Urban Intersections." Civil Engineering Dept., Purdue Univ. (July 1963).

This paper recognizes that the rating of urban highways is much more complex than the rating of rural roads. In particular, any evaluation of a major urban street must of necessity include an evaluation of the intersections on that street. This report briefly describes the development of a sufficiency rating method for such intersections based on engineering procedures. Two factors were considered to influence the ability of an intersection to serve traffic; namely, physical and traffic factors. The physical rating includes such factors as surface condition, ridability, and skid resistance, as is conventional. Also considered were intersection geometrics, curb radius for right-turning vehicles, visual restrictions, and lighting. The traffic rating used average delay per vehicle as a measure of user satisfaction with the service provided. This traffic rating is determined for each approach to the particular intersection being investigated.

SCHWAR, J. F., and PUY-HUARTE, J., "Statistical Methods in Traffic Engineering." Eng. Exper. Station, The Ohio State University (Sept. 1962).

This manual is not intended to be a textbook on statistics, as only a minimum of theory is presented. Emphasis is on the typical examples found in traffic engineering, chosen to cover those statistical methods most commonly used in the field.

"Use of Electronics in Traffic Control." Theme V, World Traffic Eng. Conf., combining 31st Ann. Meeting, Inst. of Traffic Eng. and International Sessions in Traffic Engineering (Aug. 1961).

Theme V presents seven papers from all parts of the world concerning the use of electronics in traffic control. Topics covered are vehicle guidance, simulation, computer control of traffic signals, electronic equipment, etc.

WAGNER, F. A., RUDDEN, J. B., and GERLOUGH, D. L., "Final Report, Study of the Traffic Signal System in a Portion of the District of Columbia. Volume I. Study Techniques and Results." Thompson Ramo Wooldridge, Inc. (Apr. 1963).

The purpose of the study was to "simulate signalized thoroughfares in the study area by means of Monte Carlo techniques, utilizing mathematical models on a high-speed digital computer to determine the optimum signal timing plan which will assume total delay as a measure of effectiveness." A macroscopic model of the network was formulated, and various traffic signal timing plans were tested by simulation runs on a large-scale digital computer. The model was scanned or sampled at 5-sec intervals. The distribution of the number of vehicles arriving during each 5-sec interval was assumed to be Poisson. A detailed analysis of the actual behavior of platoons on the street network under study was carried out. Numerous graphs and tables presenting platoon behavior are included. Signal timing plans have been found which suggest that delay within the network can be decreased by up to 10 percent during peak periods and up to 20 percent during off-peak periods.

WALKER, H. M., and LEV, J., *Statistical Inference.* Holt, Rinehart and Winston (1963).

This book on statistics does not presuppose any college training in mathematics, hence does not present mathematical derivations of formulas. Therefore, all concepts are introduced by the intuitive approach. Of particular interest to the project is Chapter 10, entitled "Linear Regression and Correlation," in which the authors present the results in a form suitable for machine computations.

WEBSTER, F. V., "Traffic Signal Settings." *Road Research Tech. Paper No. 39,* HMS Office (1958).

This paper presents results of research conducted by the British Road Research Laboratory into the delays to vehicles at fixed-time traffic signals. Also included is the formula for the optimum settings of such signals. The methods developed can be applied to both fixed-time and vehicle-actuated controlled intersections with some modifications. An expression is given for the average delay per vehicle in terms of both theoretical and empirical results. It is from this basic equation that the optimization concepts are developed and tables and graphs are presented of various parameters.

WILDERMUTH, B. R., "Average Vehicle Headways at Signalized Intersections." *Traffic Eng.,* pp. 14-16 (Nov. 1962).

This paper presents the results of a study made to determine the average vehicle headways at signalized intersections under different conditions. The factors which were considered to influence headways were length of phase, amount of heavy vehicles, amount of turning traffic, and direction of traffic movement. The results indicate headways in agreement with the *Highway Capacity Manual* for short green phases. Headways were found to be below 2.0 sec for phases between 35 and 45 sec (note that 2.0 sec represents 1,800 vph of green, approximating freeway measurements). Details of the analysis are presented.

YARDENI, L. A., "Vehicular Traffic Control: A Time-Space Design Model." IBM Corp. (Nov. 1964).

This paper describes a computer program which can be used for the design of traffic signal settings to produce progressions for fixed-time control. The program uses least-squares and minimax fit models to derive through-bands for given volume requirements and within the given limits of speed and cycle times.

Published reports of the

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

are available from:

Highway Research Board
National Academy of Sciences
2101 Constitution Avenue
Washington, D.C. 20418

NCHRP Report No.	*Title*
—*	A Critical Review of Literature Treating Methods of Identifying Aggregates Subject to Destructive Volume Change When Frozen in Concrete and a Proposed Program of Research—Intermediate Report (Project 4-3(2)) 81 pp. $1.80
1	Evaluation of Methods of Replacement of Deteriorated Concrete in Structures (Project 6-8) 56 pp. $2.80
2	An Introduction to Guidelines for Satellite Studies of Pavement Performance (Project 1-1) 19 pp. $1.80
2A	Guidelines for Satellite Studies of Pavement Performance 85 pp.+ 9 figs., 26 tables, 4 app. $3.00
3	Improved Criteria for Traffic Signals at Individual Intersections—Interim Report (Project 3-5) 36 pp. $1.60
4	Non-Chemical Methods of Snow and Ice Control on Highway Structures (Project 6-2) 74 pp. $3.20
5	Effects of Different Methods of Stockpiling Aggregates—Interim Report (Project 10-3) 48 pp. $2.00
6	Means of Locating and Communicating with Disabled Vehicles—Interim Report (Project 3-4) 56 pp. $3.20
7	Comparison of Different Methods of Measuring Pavement Condition—Interim Report (Project 1-2) 29 pp. $1.80
8	Synthetic Aggregates for Highway Construction (Project 4-4) 13 pp. $1.00
9	Traffic Surveillance and Means of Communicating with Drivers—Interim Report (Project 3-2) 28 pp. $1.60
10	Theoretical Analysis of Structural Behavior of Road Test Flexible Pavements (Project 1-4) 31 pp $2.80
11	Effect of Control Devices on Traffic Operations—Interim Report (Project 3-6) 107 pp. $5.80
12	Identification of Aggregates Causing Poor Concrete Performance When Frozen—Interim Report (Project 4-3(1)) 47 pp. $3.00
13	Running Cost of Motor Vehicles as Affected by Highway Design—Interim Report (Project 2-5) 43 pp. $2.80
14	Density and Moisture Content Measurements by Nuclear Methods—Interim Report (Project 10-5) 32 pp. $3.00
15	Identification of Concrete Aggregates Exhibiting Frost Susceptibility—Interim Report (Project 4-3(2)) 66 pp. $4.00
16	Protective Coatings to Prevent Deterioration of Concrete by Deicing Chemicals (Project 6-3) 21 pp. $1.60
17	Development of Guidelines for Practical and Realistic Construction Specifications (Project 10-1) 109 pp. $6.00
18	Community Consequences of Highway Improvement (Project 2-2) 37 pp. $2.80
19	Economical and Effective Deicing Agents for Use on Highway Structures (Project 6-1) 19 pp. $1.20
20	Economic Study of Roadway Lighting (Project 5-4) 77 pp. $3.20
21	Detecting Variations in Load-Carrying Capacity of Flexible Pavements (Project 1-5) 30 pp. $1.40
22	Factors Influencing Flexible Pavement Performance (Project 1-3(2)) 69 pp. $2.60
23	Methods for Reducing Corrosion of Reinforcing Steel (Project 6-4) 22 pp. $1.40
24	Urban Travel Patterns for Airports, Shopping Centers, and Industrial Plants (Project 7-1) 116 pp. $5.20
25	Potential Uses of Sonic and Ultrasonic Devices in Highway Construction (Project 10-7) 48 pp. $2.00
26	Development of Uniform Procedures for Establishing Construction Equipment Rental Rates (Project 13-1) 33 pp. $1.60
27	Physical Factors Influencing Resistance of Concrete to Deicing Agents (Project 6-5) 41 pp. $2.00
28	Surveillance Methods and Ways and Means of Communicating with Drivers (Project 3-2) 66 pp. $2.60
29	Digital-Computer-Controlled Traffic Signal System for a Small City (Project 3-2) 82 pp. $4.00

* Highway Research Board Special Report 80.

THE NATIONAL ACADEMY OF SCIENCES is a private, honorary organization of more than 700 scientists and engineers elected on the basis of outstanding contributions to knowledge. Established by a Congressional Act of Incorporation signed by President Abraham Lincoln on March 3, 1863, and supported by private and public funds, the Academy works to further science and its use for the general welfare by bringing together the most qualified individuals to deal with scientific and technological problems of broad significance.

Under the terms of its Congressional charter, the Academy is also called upon to act as an official—yet independent—adviser to the Federal Government in any matter of science and technology. This provision accounts for the close ties that have always existed between the Academy and the Government, although the Academy is not a governmental agency and its activities are not limited to those on behalf of the Government.

THE NATIONAL ACADEMY OF ENGINEERING was established on December 5, 1964. On that date the Council of the National Academy of Sciences, under the authority of its Act of Incorporation, adopted Articles of Organization bringing the National Academy of Engineering into being, independent and autonomous in its organization and the election of its members, and closely coordinated with the National Academy of Sciences in its advisory activities. The two Academies join in the furtherance of science and engineering and share the responsibility of advising the Federal Government, upon request, on any subject of science or technology.

THE NATIONAL RESEARCH COUNCIL was organized as an agency of the National Academy of Sciences in 1916, at the request of President Wilson, to enable the broad community of U. S. scientists and engineers to associate their efforts with the limited membership of the Academy in service to science and the nation. Its members, who receive their appointments from the President of the National Academy of Sciences, are drawn from academic, industrial and government organizations throughout the country. The National Research Council serves both Academies in the discharge of their responsibilities.

Supported by private and public contributions, grants, and contracts, and voluntary contributions of time and effort by several thousand of the nation's leading scientists and engineers, the Academies and their Research Council thus work to serve the national interest, to foster the sound development of science and engineering, and to promote their effective application for the benefit of society.

THE DIVISION OF ENGINEERING is one of the eight major Divisions into which the National Research Council is organized for the conduct of its work. Its membership includes representatives of the nation's leading technical societies as well as a number of members-at-large. Its Chairman is appointed by the Council of the Academy of Sciences upon nomination by the Council of the Academy of Engineering.

THE HIGHWAY RESEARCH BOARD, organized November 11, 1920, as an agency of the Division of Engineering, is a cooperative organization of the highway technologists of America operating under the auspices of the National Research Council and with the support of the several highway departments, the Bureau of Public Roads, and many other organizations interested in the development of highway transportation. The purposes of the Board are to encourage research and to provide a national clearinghouse and correlation service for research activities and information on highway administration and technology.