Control Considerations and Smooth Flow in Vehicular Traffic Nets

WILLIAM R. McSHANE, H. NATHAN YAGODA,
LOUIS J. PIGNATARO, and KENNETH W. CROWLEY,
Division of Transportation Planning, Polytechnic Institute of Brooklyn

The report presents some analysis, constraints, and problem areas in the formal structure of control (and of some particular controllers) on the signalization-assignment level of a multilevel macroscopic controller. The overall structure and philosophy of the multilevel controller itself is not presented in the present report, nor are the computer aspects considered explicitly, the digital computer being considered only as a control device for present purposes. A "smooth-flow" control philosophy is presented, and some consequent controllers for arterials and for networks are demonstrated. Some formal constraints and difficulties in structuring the control problem are treated; these include cycle length definition, predictor requirements, and the problem of, and bounds on, the closure integers. Attendant problems in estimation, computation, and stability are illustrated. An arterial controller is developed in detail; network cases are discussed.

•IN RECENT YEARS there has been a number of study projects and research presentations on digital computer control of arterials and networks. For the most part, these have been reports of specific computer-detector-signal hardware configurations utilizing a selection of predesigned stored signal plans (1, 2, 3, 4). Some are treatments of specific algorithms by which such signal plans are computed (5, 6, 7, 8). Detailed comparative studies of policies have been reported only in San Jose and Glasgow (9, 10); such studies are also to be undertaken in projects in Washington and Dallas (11, 12). No comprehensive study of control philosophy and methodology for networks on the order of the Gulf Freeway report (13) has appeared, although a paper by Ferrate (14) and a report by TRW (15) contribute significantly to filling this void.

Although there are significant questions of speed, storage, on-line control computation versus stored libraries, configurations and hierarchical systems, and so forth to be considered in any installation of digital control, the machine itself in the control loop is not considered explicitly in the present paper because it is in fact a control device and not a controller or control law per se.

It is worth noting, incidentally, that the prime advantages of digital computer control over hard-wire or analog control, other than possible cost effectiveness and central management, are (a) flexibility in revising policy, a necessity considering the existing state of the art, and (b) flexibility in future expansion of the system. Furthermore, it should be noted that computer control does not necessarily, and need not, imply closed-loop or adaptive control; several highly effective installations are essentially open-loop n-dial controllers.

CONTROL PHILOSOPHY

Control may be exercised according to several structures: multilevel, single level microscopic, and single level macroscopic. Depending on the formulation, the last alternate may be adopted to be contained within the first structure. A prime example

Paper sponsored by Committee on Theory of Traffic Flow and presented at the 49th Annual Meeting.

of the multilevel structure is the West London experiment (3) in which "strategic" decisions of a suitable policy set were made on the order of every 10 minutes, "tactical" review and adjustments were made every cycle, and "local" refinements based on local measurements were made every second or so.

The present paper is to be restricted to control in response to known or measured parameters and related problems.

Microscopic control formulations offer a satisfying formalism, particularly in terms of conventional discrete control theory. It is difficult, however, to analyze the dynamics and stability of such a formulation, or to make general statements about the coordination of intersections and the evaluation of criteria other than the one specifically employed.

Macroscopic control intervals are utilized in this presentation because they allow for formulation in terms of and comparison of standard criteria; use of the common concepts of phase, cycle, and offset; and analysis of some stability and estimation considerations.

OBJECTIVES

The objectives of the present paper are (a) to present smooth-flow control philosophy and some consequent controllers for arterials and networks; (b) to present for consideration the formal constraints and difficulties in structuring the control problem, including the necessity for certain revised definitions; and (c) to illustrate the attendant problems of estimation, computation, and stability. The control process is considered on discrete macroscopic time intervals; that is, the signalization control is set and revised on a cycle-by-cycle basis rather than on a second-by-second basis. The problem of convergence to an optimum state, equally applicable to off-line design simulators, is also presented. The structure and philosophy of the multilevel controller, as is being developed for the Dallas project, are not treated in the present problem.

THE BASIC SMOOTH-FLOW CONCEPT

The design of arterial progressions has historically meant the design of maximum bandwidth progressions, primarily because of computational feasibility. With the widespread availability of digital computers, not only did this formulation become automated and sophisticated (6, 7) but also alternate formulations were introduced as feasible (5, 8). Some utilized simulators as an integral part of the technique (17, 18).

The fact that maximum bandwidth controllers are not necessarily minimum delay policies on even one-way arterials was demonstrated by Bavarez and Newell (19) via a fluid-flow model. The prime cause of this is queuing along the arterial.

Yagoda introduced in an earlier paper (20) the use of queue-clearance or smooth-flow control on a one-way arterial, with emphasis on its use in closed-loop control. This same policy is of course applicable if the input flows and the internal turning patterns are well known, in which case it yields appropriate open-loop controllers.

Let the link offset t_i be defined as the time between the initiation of green on its 2 associated signals, downstream relative to upstream. The goal of the smooth-flow control policy is to choose a feasible set of offsets such that platoons entering any link are forced neither to decelerate nor to utilize the allocated downstream green poorly (i.e., create "slop"). This goal is functionally related to the acceleration noise criterion and to the number, but not the duration, of stops.

The desired link offset t_{di} on the ith arterial link, a one-way street segment between 2 adjacent intersections (Fig. 1), is concisely expressed in seconds as

$$t_{di} = \frac{L_i}{v_i} - \frac{Q_i}{R_i} \tag{1}$$

where

L_i = link length, ft;
v_i = desired link speed, fps;
Q_i = queue in link i at start of downstream green, vehicles; and
R_i = service rate on link i, vps.

Given that the queues Q_i are known or may be estimated, and that the prime turning activity is at the head of the queue, the design of an appropriate one-way arterial progression is simple. Designs for two-way arterials and for more general networks, which are complicated by network closure requirements, are presented later.

It is noteworthy that this control law will result in flexible, simultaneous, or "reverse" progressions as required by the queuing pattern on the arterial (Fig. 2).

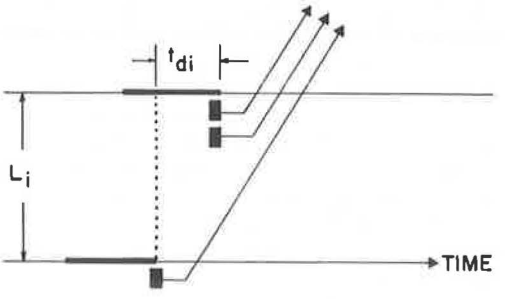


Figure 1. Smooth-flow offset.

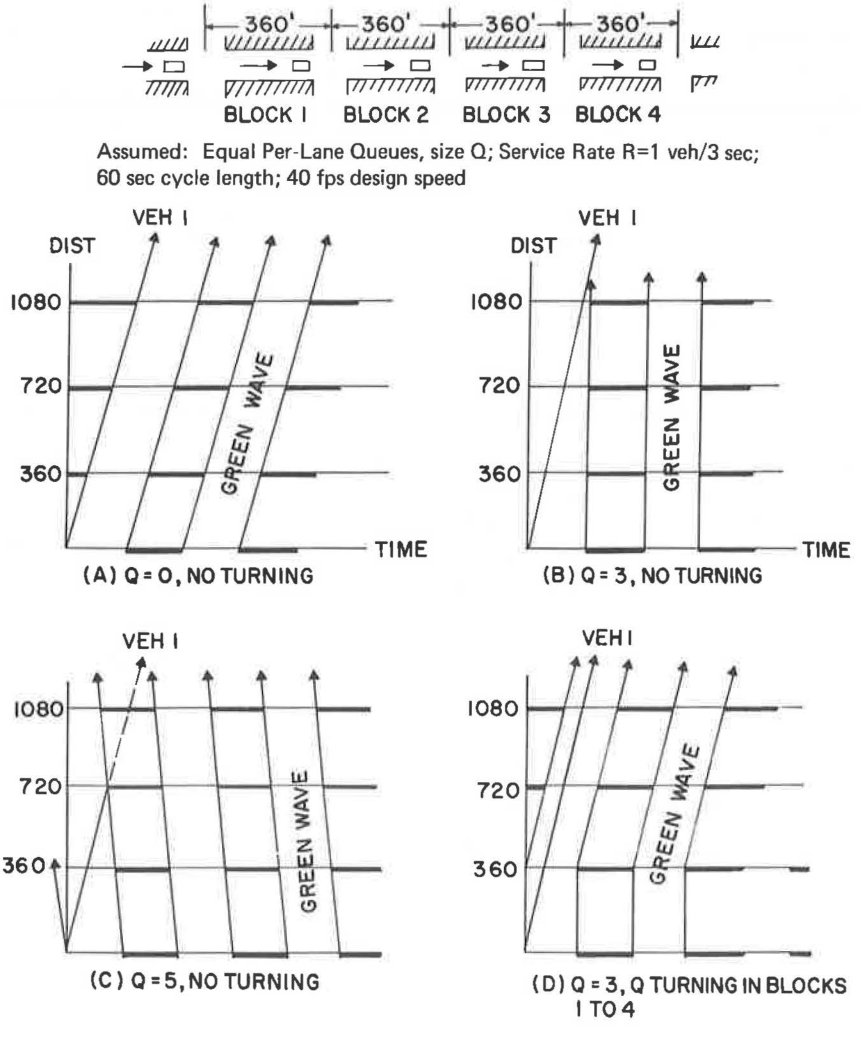


Figure 2. One-way arterial smooth-flow responses.

CLOSURE

Consider the grid shown in Figure 3A. Given the cycle length and signal splits, the offset t_4 is precisely determined if the offsets t_1, t_2, t_3 are specified. This is the crux of the network control complexity, for one must not only optimize some objective function subject to such constraints but also assure that the t_i have their minimal magnitudes in the solution, because t_i and $t_i + m_iC$ have the same physical meaning but are not mathematically equivalent in a tractable form.

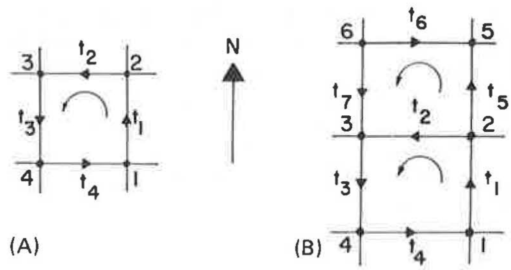


Figure 3. Closure relations on simple faces.

The closure restrictions exist physically and may be expressed mathematically for any network whose digraph (directed graph) has closed paths within it. In the development to follow, only 2 general cases are considered: (a) networks with digraphs isomorphic to an n by m digraph with quadrilateral faces (Fig. 4), and (b) the two-way arterial. Although the latter is a subcase of the former, its structure benefits from a separate formulation. Both utilize a cycle length C common to all intersections and standard green-amber-red phase sequences, with an optional all-red phase.

Let $t_{D,j}$ be the time of green initiation in direction D at intersection j. In Figure 3B, observe that with $t_{NS,1} = 0$ it follows that $t_{NS,2} = t_1$. Similarly, $t_{EW,2} = t_{NS,2} +$ (green + amber + all-red)$_{NS,2}$, which may be written

$$t_{EW,2} = t_{NS,2} + g_{NS,2}C + (a + R_d)_{NS,2}C \tag{2}$$

a set of such equations may be generated for the rest of the loop, the last of which is

$$t_{NS,1} = t_{EW,1} + g_{EW,1}C + (a + R_d)_{EW,1}C \tag{3}$$

at the original intersection. These may be reduced to

$$t_{NS,1} = (t_1 + t_2 + t_3 + t_4) + p_1C + p_2C \tag{4}$$

where

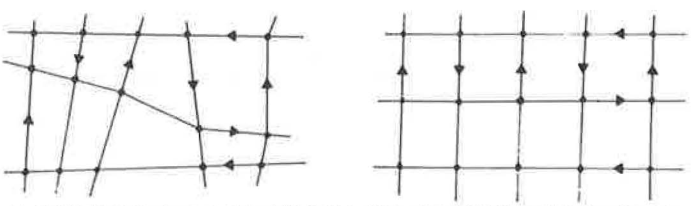
$p_1 = g_{NS,2} + g_{EW,3} + g_{NS,4} + g_{EW,1}$, and

$p_2 = (a + R_d)_{NS,2} + (a + R_d)_{EW,3} + (a + R_d)_{NS,4} + (a + R_d)_{EW,1}$.

They are referred to as the split terms. Because it was established as a reference that $t_{NS,1} = 0$, the quantity in Eq. 4 can only be a multiple of the cycle length:

$$t_1 + t_2 + t_3 + t_4 + p_1C + p_2C = nC \tag{5}$$

where n is an unknown integer. This is the basic development of a closure relation. It should be noted that the offset terms are only all positive because the links were



(A) GEOMETRIC CORRESPONDENCE (B) LATTICE REPRESENTATION

Figure 4. Physical and graphical n by m lattices.

directed with the direction of rotation in establishing the equation set. In Figure 3B, for instance, the offset from intersections 5 to 6 is $(-t_6)$; the offset sum in this case is $t_5 - t_6 + t_7 - t_8$. As a matter of convention, it is recommended that counterclockwise rotations be used on the faces and the offset terms be signed according to concurrence (+) or opposition (-) to this rotation.

For a north-south two-way arterial, the east-west links may be considered to have zero length, and the signals may be considered to merge into a single pair. Equation 5 then reduces to

$$t_1 + t_3 = nC \tag{6}$$

CONTROL STRUCTURE AND CYCLE CONVENTION

On the macroscopic control level selected for the presentation, there are 3 basic control variable sets: splits, cycle length, and offsets. It is possible to refine this level by use of submultiple cycles and special phasing, or to structure these on another level in a multilevel system, but this is not treated in the present work.

Of the 3 variable sets to be specified, one may specify a sequential selection of the 3 in some order, a simultaneous selection, or a mixed plan. Because the split allocation is inherently a response to only local intersection considerations whereas a single system cycle C and certainly the offsets have network and network coordination impact, a mixed plan is recommended. Two plans are considered: (a) simultaneous selection of cycle length and offsets, followed by split determination, and (b) sequential selection of cycle length, splits, and offsets. The split determination may be on the basis of relative flows, minimization of short-term maximum delay or lane queue, or some other criterion (20, 21).

At first inspection of some study cases, an apparent anomaly arises in the specification of varying cycle lengths and changing offsets and in the assumption of a common cycle system. Consider the simple case shown in Figure 5 of 2 adjacent signals with a 60 second common cycle, 50:50 splits, and an initial offset of zero, increased in each of 3 succeeding cycles by 5 seconds. This is achieved in the figure, but an observer at signal 2 would record cycle times of 65, 65, 65, and 60 seconds thereafter, whereas an invariant 60 seconds is observed at signal 1, an apparent violation of the common-cycle characteristic. Such difficulties could be further compounded by control outputs specifying both offset and cycle length, with an implicit multiple specification on the cycle end-point.

The apparent conflict is resolved by realizing that there must be 2 definitions of cycle length: (a) the user definition—period between green initiations at the signal of interest, and (b) the control definition—computed period of time to the next initiation of control, a period corresponding to the user cycle on an arbitrary but fixed reference signal within the system. Control is exercised synchronously throughout the system at these times, and the green phase at each signal is positioned within the cycle (and broken into 2 parts if necessary) under this definition rather than itself heralding a new cycle. While this definition is essential to the logical and formal structure of the macroscopic controller, it must be appreciated that the user sees and reacts only to the

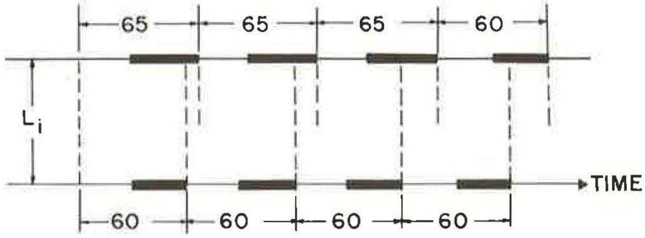


Figure 5. Anomaly in cycle length.

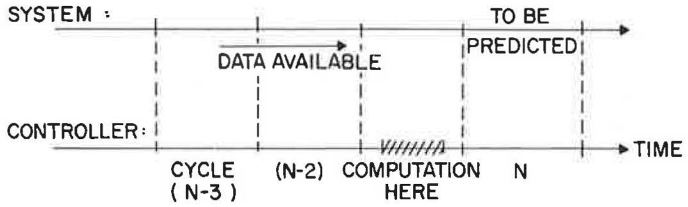


Figure 6. Need for two-step estimation.

usual succession of phases, so that bounds on the phases and other variables must be translated from one definition (user) to the other for the purpose of enforcement.

One immediate consequence of a synchronous macroscopic control structure is that the relevant system variables (i.e., those defining the state of the system for the purpose of control) must be estimated by two-step predictor techniques. This is shown in Figure 6. The control instituted at the beginning of the Nth cycle must be based on the green-initiation queues and the flows within that cycle. The processing and computation for this control instant must be done during the (N-1)th cycle, hence two-step prediction of Nth cycle state.

DELAY-RELATED OBJECTIVE FUNCTION

The existence of the closure constraints virtually precludes attainment of a complete set of desired offsets $\{t_{di}\}$ on the network of interest. Instead, a set of acutal offsets $\{t_i\}$ must be effected that is optimal according to some selected criterion and that allows for feasible vehicle design speeds v_i and cycle length C. The actual offset t_i on the ith link may be expressed in terms of the desired offset t_{di} and an offset discrepancy ϵ_i.

$$t_i = t_{di} + \epsilon_i \tag{7}$$

Common criteria for optimization are linear with constraints, absolute value, and quadratic. A linear formulation in this case would lead to a standard linear programming formulation, which is time consuming to execute and which has limited opportunity for insight. A criterion in absolute values of the offset discrepancies would generally be intractable. On the other hand, the quadratic criterion not only is tractable but also may be directly related to short-term aggregate delay for the common situations requiring network control. This is due to the observation that delay of a platoon at a signal is a parabolic function of the appropriate link offset, with minimum at some offset t_i^* (22), which is taken to correspond to t_{di} here.

The objective function or minimization criterion utilized here is therefore taken to be

$$A = \sum_{\text{links}} a_i (t_i - t_{di})^2 \tag{8}$$

where the a_i are positive weights reflecting the relative importance of links, including their vehicular content, and the efficacy of precise smooth flow on the respective links. (If the green-initiation queue Q_i exceeds the dischargeable queue for the next cycle, the arriving queue need not be timed precisely because it will stop in any case.)

THE ARTERIAL FORMULATION

Consider an N-block two-way arterial with the 2N interior links labeled as shown in Figure 7. Although there are N(N + 1)/2 distinct closure equations, only N of these are independent. Both to ensure independence and to obtain convenient bounds on the n_i as

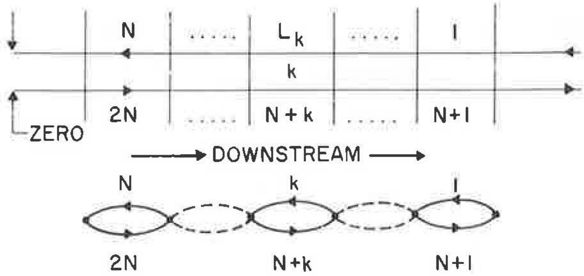


Figure 7. Conventions and equivalent graph for a two-way arterial.

discussed later, the N equations corresponding to the individual blocks and of the form of Eq. 6 are selected.

$$t_i + t_{i+N} = n_i C \qquad 1 \le i \le N$$

The formulation is facilitated by the following definitions:

1. Link i speed parameter: $w_i = 1/v_i$;
2. Block i speed parameter: $X_i = w_i + w_{i+N}$; and
3. Block i queue service parameter: $\zeta_i = Q_i/R_i + Q_{i+N}/R_{i+N}$. v_i, Q_i, and R_i have been previously defined.

The first and prime arterial case to be considered treats the classical problem of choice of an optimal cycle length C and upstream and downstream design speeds V_A and V_B respectively, subject to certain bounds on each. The constant nominal arterial design speeds yield

$$w_i = \begin{cases} 1/V_A & 1 \le i \le N \\ 1/V_B & N + 1 \le i \le 2N \end{cases}$$

and the X_i are all equal, so that the subscript may be deleted.

The formal statement of this case is as follows: Minimize the objective function

$$A = \sum_{i=1}^{2N} a_i \epsilon_i^2$$

subject to the closure relations

$$L_i X + \epsilon_i + \epsilon_{i+N} - \zeta_i = n_i C \qquad 1 \le i \le N$$

and the inequalities

$$C_1 \le C \le C_2$$

$$V_{A1} \le V_A \le V_{A2}$$

$$V_{B1} \le V_B \le V_{B2}$$

that constitute a feasible region R. The last 2 constraints yield $X_1 \le X \le X_2$.

It is significant that the objective function A is formulated in terms of offset-discrepancies, which are relevant to the system user, rather than larger loop closure discrepancy terms as have appeared in other formulations.

Assuming that the set $\{n_i\}$ is known, and for any set $\{V_A, V_B, C\}$ in the region R, one may minimize the function A to obtain

$$A = \frac{C^2}{2} \sum_{i=1}^{N} b_i \left(n_i - \frac{L_i X - \zeta_i}{C}\right)^2 \qquad (9)$$

where

$$b_i = \frac{2a_i a_{i+N}}{a_i + a_{i+N}},$$

$$\epsilon_{i+N} = \frac{a_i}{a_i + a_{i+N}} \{n_i C - L_i X + \zeta_i\}, \text{ and}$$

$$\epsilon_i = \frac{a_{i+N}}{a_i} \epsilon_{i+N}$$

so that the discrepancies within any block of the arterial are always of the same sign: Both lag, or both lead, the desired offsets.

Although it is relevant to minimize Eq. 9 with respect to speed inverses, a degeneracy results. Rather, it was decided to minimize with respect to the joint parameter X and obtain a family of (V_A, V_B) from which a suitable pair may be chosen in a second step.

The objective function A is a unimodel function of X and C for a known set $\{n_i\}$ with a unique minimum, and the constrained optimum (X^*, C^*) may be found rather easily, as shown in Figure 8. The only degenerate case occurs when the n_i are such that $n_j l_i = l_j n_i$ for all i, j, in which case there is a line $C = XL_1/n_1$ of constrained optima, which may pass through R. Even this degeneracy, however, depends only on the geometrics of the network and not on its state $\{Q_i\}$ or its weights $\{a_i\}$.

In spite of this fortuitous outcome, the arterial problem is quite complex because one must ascertain that the optimum obtained, A^*, is the minimal value for all possible sets $\{n_i\}$. Unfortunately, the minimal surface formed by the intersection of the different unimodel $A(X, C)$ for given sets $\{n_i\}$ is not itself unimodal. A simple example of this problem is shown in Figure 9.

BOUNDS ON THE n_i AND SEARCHES

Given this complication, it is of immediate interest to enumerate or at least to establish an upper bound on the feasible sets of $\{n_i\}$. Consider any random draw of N integers, with repetition permissible. These integers form a set $\{n_i\}$, and the corresponding function $A(X, C)$ has values in the feasible region R′ of the (X, C) plane that is a mapping of the feasible region R in (V_A, V_B, C) space. This does not necessarily mean, however, that this set $\{n_i\}$ has potential for yielding a physically relevant solution.

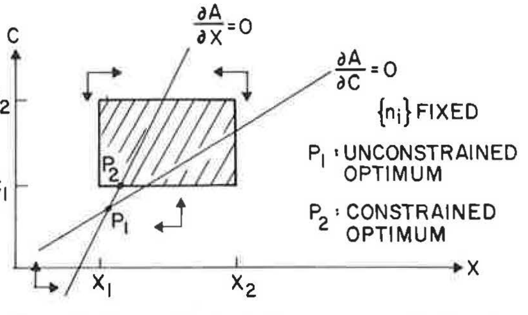


Figure 8. Constrained solution on an arterial, Case 1.

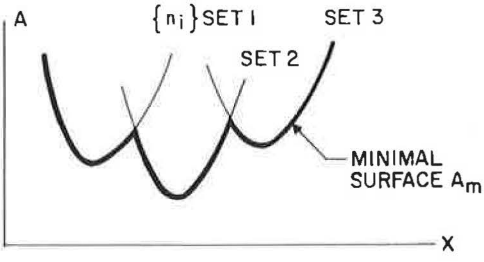


Figure 9. Example of nonunimodal surface generated by unimodal surfaces.

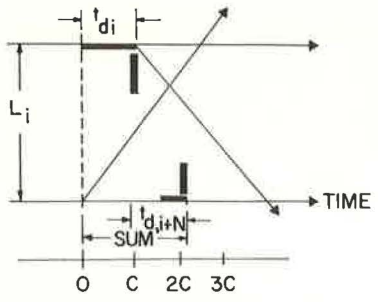


Figure 10. Bounding the integer n_i.

In Figure 10, the offsets t_{di} and $t_{d,i+N}$ are shown for a specified V_A and V_B, as is their sum. For a specified cycle length C the integer n_i should be selected to minimize the discrepancy between the sum point and the cycle multiple point. That is,

$$|(t_{di} + t_{d,i+N}) - n_iC| \leq \frac{1}{2} C \qquad (10)$$

This argument may be extended by considering the bounds on the speeds and then on the cycle length with the result that

$$\frac{L_iX_{1i} - \zeta_i}{C_R} - \frac{1}{2} \leq n_i \leq \frac{L_iX_{2i} - \zeta_i}{C_1} + \frac{1}{2} \qquad (11)$$

where R = 2 if $L_iX_{1i} - \zeta_i \leq 0$, R = 1 if $L_iX_{1i} - \zeta_i < 0$, and the subscripts 1 and 2 indicate lower and upper bounds respectively.

Let S be the collection of all $\{n_i\}$ whose elements satisfy Eq. 11. The set of feasible $\{n_i\}$ is a subset of S because certain values of, say, n_2 may preclude certain values of other $\{n_i\}$ permissible by Eq. 11. The number of elements in S, which is an upper bound on the number of feasible sets $\{n_i\}$, is simply the product of the N integer ranges obtained from Eq. 11. This information allows for a rational treatment of alternate realization techniques for the control law.

Consider the minimal surface A_m (X, C) over the region R′ formed by the intersection of the A(X, C) corresponding to the elements of S. The problem of finding the minimum on this surface, which is in fact the optimal cycle length-offset solution sought, may be attacked by at least 3 techniques: (a) complete enumeration of the constrained optima for each element of S and selection of the minimum; (b) solution for a lattice of values in the region R′ of the (X, C) plane and selection of a minimum; and (c) generation of N_2 random starting points in the region R′, unimodal searches for minima from these points, and selection of a minimum. Technique 1 is generally infeasible because of the number of elements in S.

The choice between techniques 2 and 3 depends in good part on the structure of A_m(X, C), of which little is known, and on the precision with which the absolute minima $A_m^*(X^*, C^*)$ must be ascertained. Tables 1 and 2 give typical lattice values A_m(X, C) in a feasible region R′. Table 4 gives a comparison of the average computation times for techniques 2 and 3 based on the assumption that A_m(X, C) has N_1 local minima in R′ with equal regions R_i'.

A 4- (←, →, ↑, ↓) or 8-directional search for a decrease in A_m(X, C) is conducted at each point with search steps equal to the lattice increments so that direct comparison is possible. Actually, the examples given in Tables 1 and 2 indicate that the equal R_i' assumption is conservative in favor of technique 2 because the R_i' corresponding to the absolute minimum tends to predominate (Fig. 11).

TABLE 1

VALUES OF OBJECTIVE FUNCTION, a_i = 1

C	X					
	0.04	0.05	0.06	0.07	0.08	0.09
30	5.43E2	7.64E2	1.10F3	3.81E2	4.13F2	4.48E2
35	9.23E2	5.81E2	3.42E2	2.54E2	5.03F2	5.47E2
40	1.25E3	1.05E3	6.41E2	3.23E2	3.31E2	6.49E2
45	1.55E3	1.63E3	1.13E3	6.42E2	3.32E2	4.36E2
50	1.60E3	2.08E3	1.74E3	1.12E3	1.83E3	3.69E2
55	1.60E3	2.50E3	2.46E3	1.79E3	1.08E3	5.52E2
60	1.60E3	2.85E3	3.18E3	2.62E3	1.74E3	9.85E2
65	1.60E3	2.91E3	3.66E3	3.49E3	2.57E3	1.87E3

TABLE 2

VALUES OF OBJECTIVE FUNCTION, a_i, WEIGHTED AS IN EQUATION 15, $f(Q_i) = 1$

C	X					
	0.04	0.05	0.06	0.07	0.08	0.09
30	3.64E2	1.25E2	8.83E1	2.49E2	2.86E2	3.10E2
35	6.60E2	3.29E2	1.30E2	1.28E2	3.30E2	3.66E2
40	9.07E2	6.81E2	3.10E2	1.32E2	1.81E2	4.24E2
45	1.13E3	1.11E3	6.37E2	2.95E2	1.49E2	2.49E2
50	1.17E3	1.46E3	1.10E3	9.95E2	2.75E2	1.80E2
55	1.17E3	1.76E3	1.64E3	1.04E3	5.59E2	2.70E2
60	1.17E3	1.99E3	2.15E3	1.63E3	9.79E2	5.17E2
65	1.17E3	2.04E3	2.51E3	2.29E3	1.54E3	9.22E2

The computation associated with techniques 2 and 3 is greatly simplified because there is only one feasible set $\{n_i\}$ associated with each lattice point (X, C), the elements of which are given by

$$\frac{L_iX - \zeta_i}{C} - \frac{1}{2} \leq n_i < \frac{L_iX - \zeta_i}{C} + \frac{1}{2} \qquad (12)$$

which is derivable from Eq. 11. This is simple to program in FORTRAN or APL/360, the latter of which was used to generate Tables 1 and 2.

SELECTION OF NOMINAL ARTERIAL SPEEDS

The selection of an optimal feasible pair (X*, C*) does not specify the actual offsets, although it does fix the offset discrepancies (Eq. 9). With the block speed parameter X*

TABLE 3

DATA FOR TABLES 1 AND 2

Block	L_i(ft)	ζ_i	Block	L_i(ft)	ζ_i
1	530	14	6	700	20
2	800	10	7	700	6
3	740	6	8	700	12
4	530	4	9	600	6
5	530	0	10	530	6

Notes: 1. Minima circled.
2. E3 ⟶ 10^3.
3. $Q_i = Q_{i+N}$.

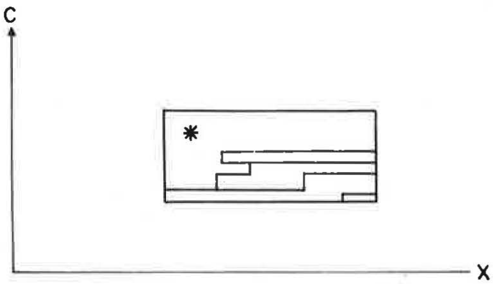


Figure 11. Typical set of R_i'.

TABLE 4

COMPARISON OF TWO SEARCH TECHNIQUES

Case	Technique	Est N_1	Number of Search Moves	Prob Spec on Failure to Find Absolute Minimum		
				0.01	0.05	0.10
1	Random start	10	4	316	209	159
2	Random start	4	8	221	143	117
3	Random start	2	4	112	80	64
4	Random start	2	8	127	91	73
5	Lattice of values	–	–	⟵ 153 ⟶		

Note: 17 by 9 lattice utilized in the (X, C) plane. All table entries normalized to single-point computation time.

determined, and with the relative precision in the 2 stream directions determined by the $\{a_i\}$, the actual assignment of values to V_A and V_B may be achieved by a second criterion with constraints. A common criterion for this purpose is one involving the speed-density relation in each direction.

OTHER ARTERIAL CASES

The general development presented here may be easily adapted to treat other cases: (a) Case 2—minimization of the objective function by selection of design speeds V_A and V_B but with a specified cycle length C; (b) Case 3—minimization of a weighted quadratic objective function in the discrepancies of speed parameters from a known desired set $\{\hat{w}_i\}$

$$A_1 = \sum_{i=1}^{2N} a_i\{w_i - \hat{w}_i\}^2 \tag{13}$$

with a specified cycle length assumed; and (c) Case 4—minimization of an objective function similar to Eq. 13 but directly in terms of speeds rather than speed parameters.

Case 2 is essentially a subcase of the prime case treated in detail earlier; the random-search technique, however, is not generally merited over the lattice technique because of reduced dimensionality.

The fact that a design parameter $\hat{w}_i$ or $\hat{v}_i$ is assumed to be known or estimated for each link in Cases 3 and 4 allows the arterial design to be decomposed into individual block designs because the commonality of the unknowns V_A and V_B is lacking and the structure of the block closures do not provide interblock linking. Cases 3 and 4 are treated without speed constraints because of the local adaptability allowed by this decomposition. The formulation in Case 3 not only allows the variable design speed arterial to be treated but also is conveniently linear. Case 4 is also a treatment of variable design speeds, but the optimization equations are nonlinear (21). The merit of one over the other is dependent on an evaluation of speed or the speed parameter as the more relevant variable to optimize; the latter is directly indicative of offset and thus the adjustment or maneuver period a driver faces as he approaches the signal.

STABILITY AND POLICY ENFORCEMENT

Little has appeared in the literature on the analysis of controllers for local and global (i.e., system) stability. In particular, the problems of oscillation of local queue sizes, of employment of a degrading objective function or estimator, or of tracking an optimum and of switching among policies are virtually untreated in the literature.

Some analysis has been done by one of the authors (21) of 2 link stability problems: (a) the interaction of discrepancy, internal queues, and objective function value in Cases 3 and 4, and (b) the effects of errors in speed-density estimators. Conditions for potential local instabilities are derived; these potentials are realized in near-saturation flows.

Speed or speed parameter discrepancies from the desired $\hat{v}_i$ or $\hat{w}_i$ within any block tend to drive the block queue-service parameter ζ_i to a value such that the objective function value is decreased (21). Under sustaining flows, this leads by iteration to an optimal value of the objective function by by error-closure or negative feedback. It also leads, by this link analysis, to a stable set $\{\zeta_i\}$ of internal queues.

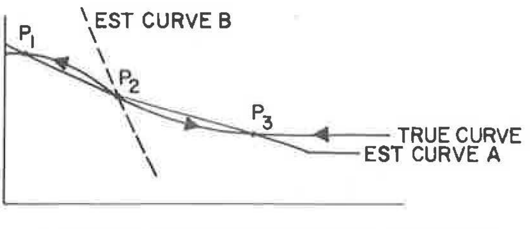


Figure 12. Stability and speed-density estimates.

The second link problem treated indicates that not all intercepts of true and estimated speed-density link relations are statically stable equilibria and that, even at the statically stable intercepts, the difference in slopes of the 2 curves determine dynamic equilibrium. For example, points P_1 and P_3 of estimator A (Fig. 12) are statically stable by this analysis while P_2 is a potential saddlepoint. Although P_2 of estimator B is statically stable, the large slope difference makes dynamic stability unlikely because of the large $\hat{v}$ with resultant large offset goal changes for small queue changes.

On the system or global level, there appears to be no satisfactory presentation anywhere in the literature. Certain control precautions, however, may well be noted. First, considering the fact that a particular offset need not be smooth-flow if the green-initiation queue Q_i exceeds what may be discharged (i.e., Q_{di}) in the next cycle, one may establish

$$a_i = \begin{cases} f(Q_i) & Q_i < Q_{Di} \\ 0.001 & Q_i \geq Q_{Di} \end{cases} \tag{14}$$

where the nonzero quantity precludes programming difficulties in control algorithms. This, however, is not recommended because local variations of Q_i around Q_{di} may cause significant value-switching in some a_i, with similar impact in the control outputs possible. Rather, a tapered form such as

$$a_i = f(Q_{Di}) \left\{\frac{Q_{Di}}{Q_i}\right\}^2 \quad Q_i \geq Q_{Di} \tag{15}$$

is recommended. Second, it must be recognized that even with tapered weights and smoothed input estimates, it is possible for the control output to switch discontinuously. This is illustrated by the data given in Table 1; if a Case 2 controller is assumed with a specified change from C = 60 to C = 65 seconds, the block speed parameter switches from X = 0.09 to 0.04 $(\text{fps})^{-1}$. As a result, the enforced offsets would radically change in a single cycle and the phasing as it appears to the user would be destroyed.

In addition to being unsatisfactory from consideration of the user, such situations have the necessary elements for a system-wide instability; that is, a discontinuous control is enforced on a system that has granularity and noise in its output and that furthermore is in a state near a switching boundary. Moreover, such switching may uncover the transport lags on long links, which are somewhat absorbed by smoothed macroscopic control.

It is therefore necessary not only to smooth the control input, which is done for noise filtering and prediction, but also to enforce only small variations at the control output. This may be done by smoothing the output of the controller or by subjecting the output variations to a saturation or by constricting the single-step feasible region R' to percentage changes of the most recent control values. This last approach, however, may fail to "follow" (or track) the true optimum well over long periods and would require updating by another technique.

ESTIMATION AND DETECTION

One problem of prime importance is the accurate two-step prediction of the state of the arterial or network, particularly the green-initiation queues Q_i. Two techniques are available: (a) measurement of the local Q_i of interest over the first through (Nth - 2) cycle and the estimation of $Q_i(N)$ on the basis of a stochastic process model for that local queue, and (b) measurement of all Q_i within a region of influence, also including flow and turning action as necessary, all through the (Nth - 2) cycle, and estimation of $Q_i(N)$ on the basis of transport of these quantitites through the 2 cycles via a simulator or analytic algorithm. Although the second approach offers significant complexities in formulation, measurement, and execution, preliminary analysis indicates that the first approach is not fruitful. This is reenforced by a recent study (15) indicating that relevant measures of effectiveness are highly sensitive to queue measurement.

THE n BY m NETWORK

The network formulation is significantly more complex than its interesting subcase, the two-way arterial; this is due to the interlocking nature of the closure relations.

Even the highly structured n by m lattice of one-way links is extremely complex, so that this section is intended to detail that complexity and to present a potentially fruitful approach rather than a completed algorithm.

An n by m lattice network with no zero-length links has N_L internal links (those links connecting 2 intersections in the defined network), N_a one-way arterials, and N_c simple faces (in this case, the area enclosed by the closure of 4 links), where

$$\left.\begin{aligned} N_L &= 2nm - (n + m) \\ N_a &= n + m \\ N_c &= (n - 1)(m - 1) \end{aligned}\right\} \quad (16)$$

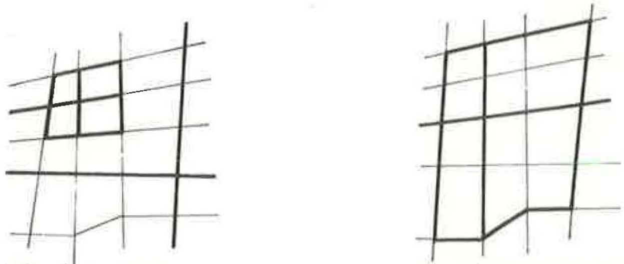
The number of closure equations is also N_c, and it is recommended that these be written on the simple faces with a consistent rotation convention, as shown in Figure 4.

Two formulations have been considered by the authors: (a) the selection of an optimal set of constant arterial speed parameters $\{w_\gamma | \gamma = A, B, \ldots\}$, one for each included arterial, with inequality constraints specified, and (b) the selection of an optimal set of arterial link discrepancies, given estimates $\{\hat{w}_i\}$ of the desired speed parameters. The first utilizes the cycle length C as a control variable; the second assumes that the cycle is predetermined.

The first formulation is rather untenable, although certain specialized cases yield interesting results. In particular, a number of small lattice networks generally have unconstrained solutions that zero the objective function while maintaining several degrees of freedom. This is true whenever the number of arterial speeds $N_a + 1$ (i.e., the cycle length, $N_a + 1 = N_d$) exceeds the number of constraining equalities N_c. Moreover, the 3 by 6 and 4 by 4 lattices have unique solutions that zero the objective function because $N_d = N_c$. Although the solutions may not be in the feasible region, they provide a basis for analysis of structural effects.

The analysis for nonzero objective function A is complicated in either formulation, however, by 2 principal factors: (a) the need to utilize the partial derivatives of A with respect to the independent control variables, resulting in a matrix whose terms involve the L_i, b_i, and n_i, and (b) the ensuing difficulty in establishing conditions for degeneracy and consistency. These problems become more pronounced with increasing lattice size, and the prospect of a generalized analysis correspondingly diminish. At the same time, however, the increase in the dimensionality of the solution space with lattice size dictates that the generalized properties be utilized to increase the efficiency of each solution.

Aside from these considerations, the merit of maintaining the constant arterial-speed assumption in a large network is questionable: Traffic levels tend to be disparate along individual arterials, and regions of similar local activity tend to become defined. Based on a consideration of these factors, as well as the specific difficulties noted earlier, 3 alternates are proposed: (a) decompose the network into a set of disconnected subnet-



(A) DECOMPOSITION TO SUBNETWORKS AND TREES

(B) DEVELOPMENT OF A MAJOR LATTICE

Figure 13. Alternates for treating large lattices.

works and trees on the basis of certain criteria and judgments to be specified, (b) establish a major lattice within the network and several secondary or minor lattices (bounded by segments of the major lattice) and solve the 2 divisions sequentially, and (c) employ a formulation based on minimizing local speed or offset discrepancies.

Alternates 1 and 2 are shown in Figure 13. Alternate 3 is treated by the second formulation cited earlier. It has been shown (21) that there is always a nonsingular solution for the speed parameter set $\{w_i\}$, expressible in terms of the closure integers $\{n_i\}$. Although this formulation is similar in several respects to Case 3 for the two-way arterial, no decomposition is possible. The principal difficulty is thus once again the selection of a certain set $\{n_i\}$.

CONTROL PERSPECTIVE

Problems and techniques related to macroscopic synchronous control of traffic on arterials and on networks have been presented and treated in varying detail, and certain controllers based on the queue-clearance or smooth-flow policy have been developed. One related point arises: the use of macroscopic control as a self-contained policy as opposed to as part of a multilevel system.

It should be noted that the macroscopic controllers developed utilize as control inputs the (estimated) link flows, queues, speeds, and, depending on the estimation model, turning rates; they are not explicitly responsive to origin-destination patterns in the network or to the equivalent flow paths or to the impeding onset of well-known usage level changes, such as rushhour.

In addition to these factors, which imply a need for a higher policy-oriented level of control, there is another consideration. As shown in Figure 14, the control state is dependent on measurement of the system or process output, without any error comparison to system input. Generation of objective function values is meant to emulate the measures of effectiveness (MOE).

It is not implausible, however, that even a fixed input vector $\underline{f}$ might yield 2 different final outputs $\underline{v}$ even under control, and result in different control outputs $\underline{p}$. This can occur because of randomness within the system, causing a $\delta(t_i)$ separation at time t_1 in the previously identical $\underline{v}(t)$ and resulting in 2 different trends in control output $\underline{p}$ because of the switching nature of the control components. The system outputs $\underline{v}_1$ and $\underline{v}_2$ may each be locally stable: This analysis treats absolute minima only for the function A(X, C) or such with the system output $\underline{v}$ fixed at any value. No theory currently exists to treat this final-state problem. One signal plan algorithm uses randomization to circumvent this difficulty (17).

Considering this discussion, one may argue that an installation utilizing on-line computation of its algorithm and employing the output state $\underline{v}$ of the physical system would need a background simulator that would run trial cases to validate that the on-line state $\underline{v}$ was optimal. This is another area, one with direct implication for on-line computation versus stored library, which has no existing theoretic treatment.

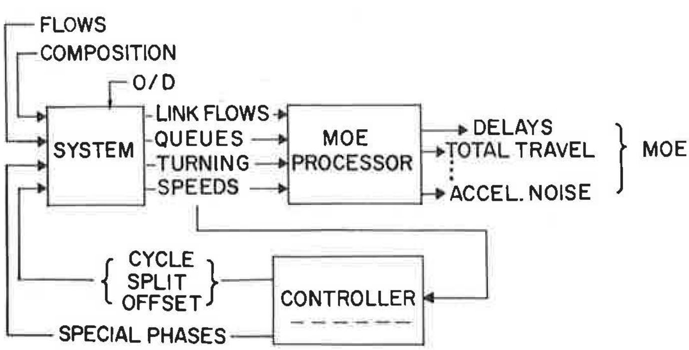


Figure 14. Control process.

Based on these considerations, it does not seem that the macroscopic controller would suffice without a higher level of control (which may involve human operation within the control loop) to preplan the response to well-known load patterns and to supervise the selection and validation of optimal system and control states. A lower level of control may be used for local special phasing, intersection blockage, or excess green reassignment; this phase is being developed for the Dallas corridor project.

FURTHER DEVELOPMENT

Further detail and exposition on several of the topics treated in this paper are available (20, 21). Related work is presently under way both as doctoral dissertation studies and as contract work on the Dallas Corridor Surveillance and Control Project for Texas Transportation Institute.

REFERENCES

1. Wilshire, R. L. The Benefits of Computer Traffic Control. Traffic Engineering, April 1969.
2. Hewton, J. T. Metropolitan Toronto Traffic Surveillance and Control System. Civil Engineering, Feb. 1969.
3. Cobbe, B. M. Traffic Control for West London. Electronics and Power, April 1967.
4. Karagheuzoff, T. T. A Perspective on Electronics in Traffic Control. Presented at IEEE: NEREM, Boston, Nov. 6, 1969.
5. Yardeni, L. A. Algorithms for Traffic-Signal Control. IBM Systems Jour., Vol. 4, 1965, pp. 148-161.
6. Morgan, J. T., and Little, J. D. Synchronizing Traffic Signals for Maximal Bandwidth. Operations Research, Vol. 12, 1964, pp. 896-912.
7. Johnson, J. Optimum Control of an Unsaturated Artery. IBM.
8. Chang, A. Synchronization of Traffic Signals in Grid Networks. IBM Jour. of Research and Development, July 1967.
9. San Jose Traffic Control Project, Final Report, 1967.
10. Hillier, J. A., and Holroyd, J. The Glasgow Experiment in Area Traffic Control. Traffic Engineering, Oct. 1969.
11. Gordon, R. L. Urban Traffic Control. Sperry Rand Engineering Review, Vol. 22, No. 1, 1969.
12. McCasland, W. R., and Carvell, J. D. Optimizing Flow in an Urban Freeway Corridor. Texas Transportation Institute, Annual Report Fiscal Year 1968-69.
13. Gap Acceptance and Traffic Interaction in the Freeway Merging Process—Phase II. Texas Transportation Institute, Final Report, 1969.
14. Ferrate, G. A. Requirements of an Advanced Hierarchy System, Using Sub-Masters, For Computer Area Traffic Control. Presented at the IFAC Conf., Haifa, Israel, 1967.
15. System Analysis Methodology in Urban Traffic Control Systems. TRW Systems Group, Rept. 11644-H014-R0-00, June 1969.
16. Busacker, R., and Saaty, T. L. Finite Graphs and Networks. McGraw-Hill, New York, 1965.
17. Irwin, N. A. The Toronto Computer-Controlled Traffic Signal System. In Traffic Control, Theory and Instrumentation, Plenum Press, New York, 1965.
18. Robertson, D. I. Transyt: A Traffic Network Study Tool. Road Research Laboratory, Crowthorne, Berkshire, England, RRL Rept. LR 253, 1969.
19. Bavarez, E., and Newell, G. F. Traffic Signal Synchronization on a One-Way Street. Transportation Science, Vol. 1, No. 2, 1967, pp. 55-73.
20. Yagoda, H. N. The Control of Arterial Street Traffic. Proc. IFAC Conf., Haifa, Israel, 1967.
21. McShane, W. R. The Control of Vehicular Traffic on Urban Arterials and Networks. Polytechnic Institute of Brooklyn, PhD dissertation, June 1968.
22. Hillier, J. A., and Rothery, R. The Synchronization of Traffic Signals for Minimum Delay. Transportation Science, Vol. 1, No. 2, 1967, pp. 81-94.