Algorithms for the design of traffic-signal progressions for fixed-time control are described.

Least-squares and minimax fits are used to derive solutions for given volume requirements within specified limits of speed and cycle time.

The algorithms have been programmed for processing on a digital computer, thus reducing the initial design time considerably and leading to solutions that are superior to manually derived designs.

Algorithms for traffic-signal control

by L. A. Yardeni

Control of vehicular traffic presents a problem of ever-increasing severity, especially along arteries with large numbers of traffic signals. Many attempts have been made to set the timing of the signals for satisfactory traffic flow. However, most of these designs for fixed-time control require a manual analysis of the traffic situation. Such designs are time-consuming and do not always lead to the best possible solution.

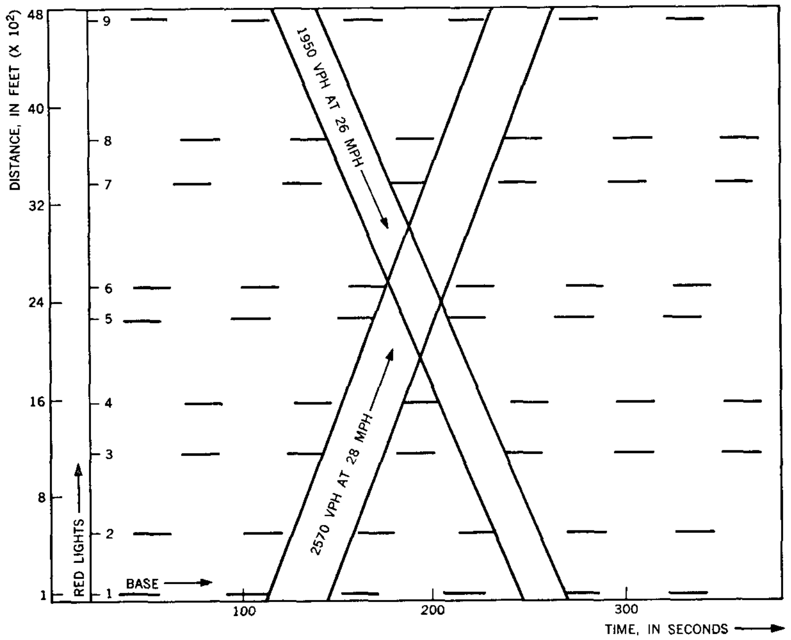
This paper presents two algorithms that have been programmed for processing on a digital computer. The initial work in developing good traffic-signal progressions is reduced from days to minutes, and solutions superior to those derived manually are realized.[1]

the problem

Major traffic arteries are characterized by a string of signalized intersections and relatively heavy vehicular flow in either or both directions during at least some hours of the day. Ideally, any one vehicle or group of vehicles should be allowed to proceed along the artery at a suitable speed without stopping or slowing down due to red signal intervals. Although such an ideal situation can be achieved rarely for two-way streets, it can be approached by properly setting all the traffic signals along the artery.

The problem consists of finding the correct timing for each signal in order to arrive at the best overall solution. Presently, city traffic engineers find a solution (which is not necessarily optimal) by drawing time/space diagrams, as shown in Figure 1. The vertical distances between the horizontal lines indicate the true-scale

Figure 1 Time/space diagram



distances of the signalized intersections. The lines themselves represent the red-signal interval times in both directions, whereas the blanks between reds are the green (plus amber) interval times, called *green split* (expressed in seconds) or *green ratio* (percentage of the entire signal cycle). The two pairs of parallel diagonal lines enclose the *through-bands* in opposite directions; their slope represents the speed a vehicle must maintain in order to stay as part of a group, or *platoon*, of vehicles moving along the major street without signal interference.

The essential input information for a time/space diagram is limited to the distances between the intersections in the system and the green ratios for these signals. For the conventional drawing board design, various sets of speeds and cycle times are used with green splits precomputed from averaged volume data. The number of such input sets that can be tried for the best solution is limited by the time required to draw and evaluate the results. The algorithms described in this paper overcome this limitation because they can be programmed for digital computer application.

Usually, an attempt is made to design for higher speeds within the legal speed limits. Shorter cycle times, such as 60 seconds or less, are also preferred, because they provide shorter average waiting time for vehicles stopped at a red signal. But in view of acceleration losses due to signal switching, shorter cycles reduce the feasible volume (vehicles per hour). The algorithms described here take into account these conflicting requirements: they allow a computer to select the highest feasible speeds and the lowest cycle time within given ranges. The computer accepts volume data for each intersection and computes proportional green ratios in preparation for the design computations. It also considers input data for the free-flow volume versus speed relations in computing actual flow capabilities of the through-bands.

objectives

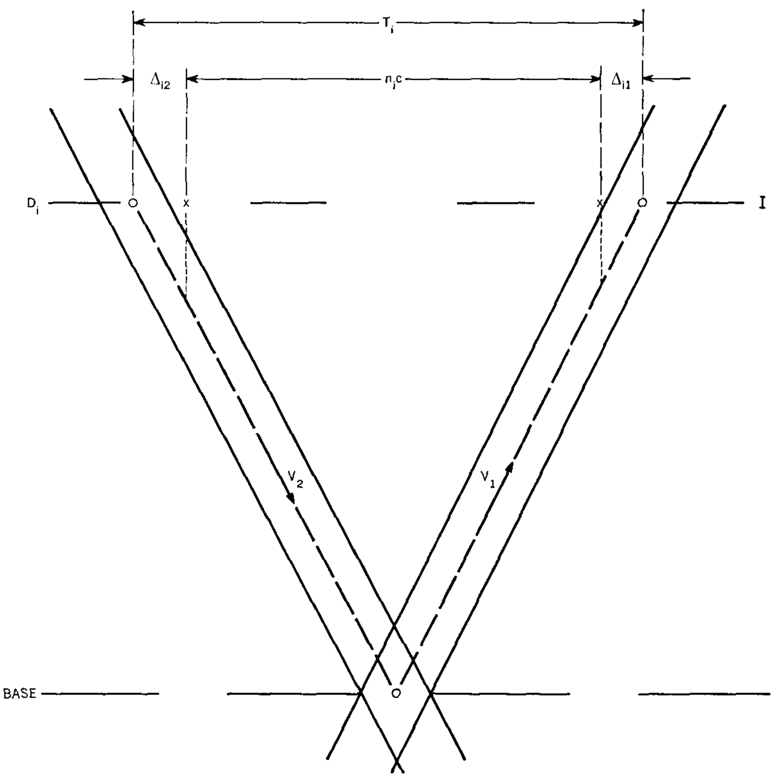
The quality of the design may depend on the criterion of optimality or measure of effectiveness chosen. Typical design objectives may be defined as follows: (1) minimize red-phase interference in the through-band, (2) maximize through-band widths, (3) maximize some criteria of flow quality, or (4) select a maximum speed and minimum cycle combination that meets flow demands. The last design objective listed is used in the approach of this paper. It is felt that this objective is superior to other optimization criteria examined, because it addresses itself explicitly to meeting traffic throughput requirements.

The algorithms

initial base bands

From a set of intersections along an artery, the intersection with the lowest green ratio is selected as a base. All other intersections may now be considered independently in relation to this base. Through-bands enclosing an entire green split of the base intersection are initially the widest possible bands for a given set of green ratios. Figure 2 shows the relation of any other intersection to the base.

Figure 2 Midpoint relations



Consider any intersection I at a distance D_i from the base. Assuming any pair of speeds V_1 and V_2, we may draw them at their proper slopes through the midpoint of the base's green interval. Transposing this point along both directions to I, we obtain the *turn-around time* T_i which is given by the sum of travel times in both directions:

$$T_i = \frac{D_i}{V_1} + \frac{D_i}{V_2} = D_i\left(\frac{1}{V_1} + \frac{1}{V_2}\right).$$

Also,

$$T_i = \frac{2D_i}{V}, \tag{1}$$

where

$$V = \frac{2V_1V_2}{V_1 + V_2}.$$

To assure that the signal timing for intersection I does not interfere with the initial base bands, it is required that:

$$T_i = c_i n_i, \tag{2}$$

or

$$\frac{T_i}{c_i} = n_i,$$

where n_i is an integer that yields a cycle c_i within a given allowed range (e.g., 50 to 100 seconds). This is an ideal cycle time which would eliminate any red phase band interference for intersection I.

In a real system, it is seldom feasible to select a system cycle time c that properly satisfies the above requirement simultaneously for all intersections.

For system cycle c, the deviation Δ_i from the ideal turn-around time for intersection I, T_i, is then $\Delta_i = \Delta_{i1} + \Delta_{i2} = T_i - cn_i$.

The optimization objective is now to minimize these deviations for all intersections simultaneously. To do this, the conventional *least-squares fit* is applied.

least-squares linear fit

The criterion function A is now defined as:

$$A = \sum_i (T_i - cn_i)^2,$$

or, by using Equation 2,

$$A = \sum_i (c_i - c)^2 n_i^2. \tag{3}$$

Minimization with respect to c yields the best cycle time for a given pair of speeds, V_1 and V_2, or a mean speed V.

Some manipulation of variables gives other equivalent forms of this function as follows: Replacing T_i from Equation 1,

$$A = \sum_i \left(\frac{2}{V} D_i - cn_i\right)^2 = \frac{4}{V^2} \sum_i (D_i - \tfrac{1}{2}cVn_i)^2$$

$$= \frac{4}{V^2} \sum_i (D_i - Kn_i)^2, \qquad (4)$$

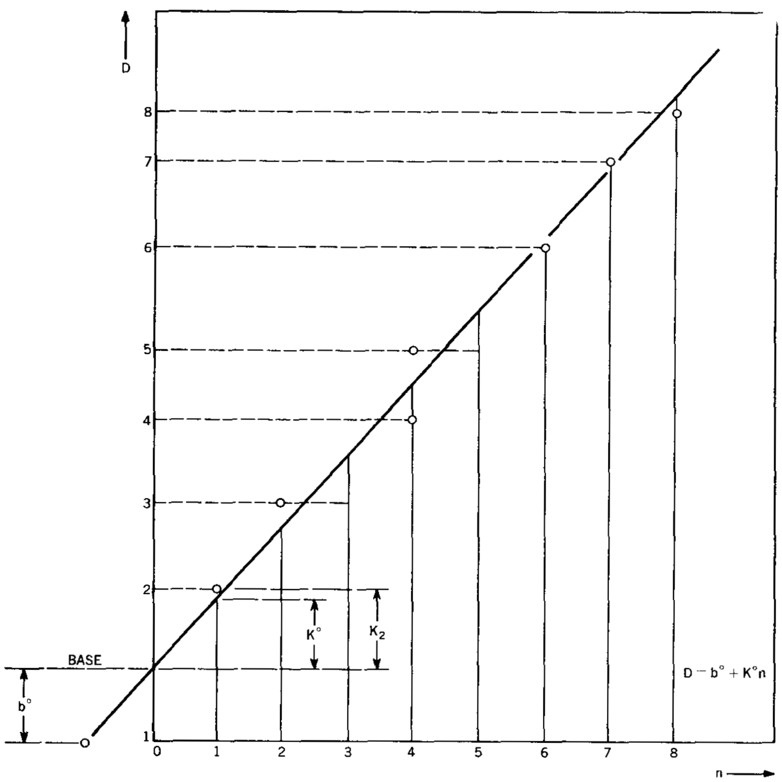
where $K = cV/2$ is a system constant, the *space-periodicity constant*, to be discussed further.

Equations 3 or 4 may be used to calculate an optimal cycle length (c^o) for given speeds, an optimal average speed (V^o) for given cycle lengths, or an optimal system constant (K^o).

Any pair of c and V satisfying the condition $K^o = cV/2$ is equally optimal with respect to the criteria of minimum A. As shown later, the actual selection of c and V values is governed by other considerations, whereas the function A is defined in terms of the systems constant K only.

Before differentiation and evaluation of a formula for optimal values K^o of K, the function A is first generalized. According to Equation 4 and Figure 3, K^o is the slope of a best fit through the discrete points relating D_i and n_i.

Figure 3 Space-periodicity fit



The first obvious generalization is to provide for an intercept, or a base line, which does not necessarily coincide with the location of any real intersection, as follows:

$$A = \frac{4}{V^2} \sum_i (D_i - b - Kn_i)^2. \tag{5}$$

Both K and b are expressed in units of distance, i.e., feet; b is the distance of an imaginary base intersection from the first intersection; K is the system's space periodicity constant. In the present formulation, K depends primarily on the given fixed distances $(D_i - b)$; it may possibly have more than one optimal value if one or more intersections are assigned different values of n_i. For some intersections, its relative position to the best-fit line of slope K may be very close to flipping from one value of n_i to the next; e.g., intersection 5 in Figure 3 may be assigned $n_5 = 4$, but would be assigned $n_5 = 5$ for a very slight change in K. This would yield another value of K^o. To select the best K^o, we can plug the different K^o back into Equation 5 until we find the lowest value of A, use some other goodness-of-fit test, or test for expected bandwidth.

algorithm refinements

To improve the fitting even further, we modify the algorithm to consider only relevant deviations. This is done by taking into account all differences in green splits.

If the green interval of intersection I in Figure 2 is larger than the base green interval, it is not necessary to have $\Delta_{i1} = \Delta_{i2} = 0$ for noninterference in the through-bands. This consideration allows relaxation of the requirement expressed by Equation 2. The modified formula is of the form

$$T_i = c_i n_i (1 \pm g_i),$$

where

$$0 \leq g_i < 0.5,$$

g_i being an appropriate expression for the difference between the green ratios of the intersection I and the base. The sign (+ or −), depending on the occurrence of a *lead-trail* or *trail-lead* fit (as explained later), tends to decrease the derivations in the least-squares fit, as shown in Figure 4.

Redefining our integers n_i as modified integers $m_i = n_i(1 \pm g_i)$ we obtain:

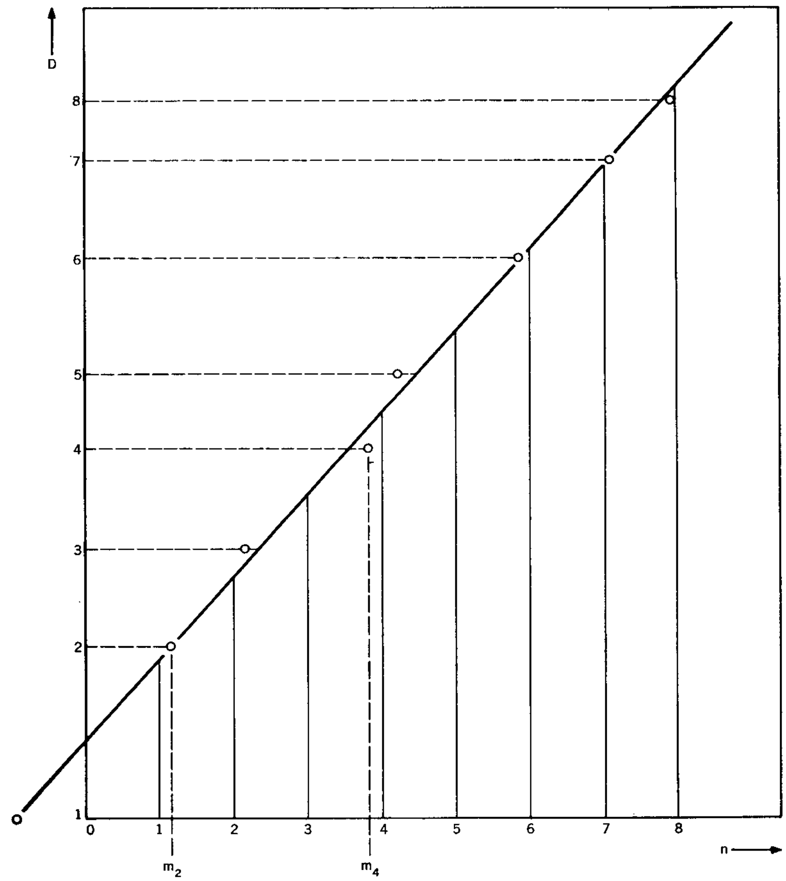
$$A = \frac{4}{V^2} \sum_i (D_i - b - Km_i)^2. \tag{6}$$

So far, this criterion function is essentially a sum of squares of rather rigidly defined deviations. It lends itself readily to the customary least-squares fit. Prior to performing such a fit, an additional generalization is now introduced: the weighting of the deviations by a weighting coefficient a_i so that

$$A = \frac{4}{V^2} \sum_i (D_i - b - Km_i)^2 a_i^2. \tag{7}$$

The introduction of this factor obviously renders considerable flexibility to our formulation and can provide the decision maker with a range of distinct models. Such models depend on the method

Figure 4 Space-periodicity fit with modified n_i, m_i



used for the selection of a_i's. For example, consider the following cases:

(1) All $a_i = 1$ (obviously, this trivial case yields Equation 6).

(2) $a_i = \max_L \{F_{iL}\}$,

where F_{iL} is later defined by Equation 9 as a function of required volume and green ratio for intersection I in direction L. In this model, the deviations are thus weighted by coefficients that provide a ranking of the intersections with respect to their *degree of criticality*. A critical intersection is loosely considered as one that imposes design restrictions on the system.

(3) Modification of a_i's as a result of an iteration procedure. For example, starting with all $a_i = 1$, design through-bands and determine the band limiting intersections, and then assign values $a_i > 1$ to these intersections and redesign. This may be repeated any desired number of times. Finally select the design with maximum band widths.

Differentiating Equation 7 with respect to K and b, and setting $\partial A/\partial K$ and $\partial A/\partial b$ equal to zero, K° and b° are derived:

$$K^{\circ} = \frac{\sum a_i^2 \sum a_i^2 D_i m_i - \sum a_i^2 D_i \sum a_i^2 m_i}{\sum a_i^2 \sum a_i^2 m_i^2 - (\sum a_i^2 m_i)^2}$$

and

$$b^{\circ} = \frac{1}{\sum a_i^2} \left(\sum a_i^2 D_i - K^{\circ} \sum a_i^2 m_i\right),$$

where the summations are taken over all intersections; D_i is the distance of any intersection I from the first intersection or, more generally, from any reference point in the system if the m_i's are derived and b is measured with respect to that same reference point; m_i is the modified integer; and a_i is a properly defined weighting coefficient for intersection I.

expected bandwidth test

As implied earlier, different values of K° and b° may be obtained by assigning different values to the set $\{m_i\}$. A simple method provides the selection of *all* feasible sets within the given ranges of speeds and cycle times. Clearly, each set of $\{K^{\circ}, b^{\circ}\}$ is an optimum with respect to a distinct set $\{m_i\}$. Some possible discrimination and selection criteria for the best $\{K^{\circ}, b^{\circ}\}$ have been mentioned. Another possible criterion is developed next.

Just as we defined an ideal cycle time (c_i) for intersection I, an ideal constant (K_i) can be considered as follows:

$$K_i = \frac{D_i - b^{\circ}}{m_i},$$

so that any one deviation in the least-squares fit is:

$$d_i = (K_i - K^{\circ})m_i.$$

This expression, when used in its absolute form, is useful in obtaining an estimate of expectable through-band widths, and further in developing another selection criteria for K°.

First, the largest deviation is found:

$$d_w = \max_i \{d_i\} = \max_i \{|K_i - K^{\circ}|\, m_i\},$$

from which the worst K_i, say K_w, is obtained.

Next it is recognized that the minimum reduction in the initially feasible through-band width is approximately:

$$\Delta B \cong \frac{2}{V} |K_w - K^{\circ}| (1 - g_w) = \frac{2}{Vc} \cdot \Delta K_w \cdot (1 - g_w) \cdot c$$

$$= \frac{\Delta K_w}{K^{\circ}} c(1 - g_w)$$

where g_w is the absolute difference in green ratios as earlier defined.

Now let G_{bL} equal the green interval (in seconds) of the imaginary base in direction L. This is the initially feasible band width. Then the maximum expected band width is given by:

$$B_{\max, L} = G_{bL} - \Delta B_L = c \cdot \mathcal{G}_{bL} - k_r \cdot c(1 - g_{wL}) \quad (8)$$

where

$$\mathcal{G}_{bL} = \frac{G_{bL}}{c}$$

is the green ratio for the base in direction L, and

$$k_r = \frac{\Delta K_w}{K^o}$$

is the relatively worst deviation from the system's periodicity constant.

If we set the right-hand side of Equation 8 equal to zero, we obtain the conditions for through-band feasibility, i.e.,

$$0 \leq k_r < \frac{\mathcal{G}_{bL}}{1 - g_w}.$$

Thus, an estimate of through-band feasibility and relative magnitude can be obtained from fixed distance and variable flow information only, independent of cycle or interval times and speeds. However, this assumes that green ratios are selected in some fixed relation to volume ratios.

minimax linear fit

An alternative algorithm is based on the criterion function

$$\{\delta_i\} = \{D_i - b - Km_i\}$$

as shown in Equation 6. In this approach, K^o and b^o are evaluated and alternative values selected by a minimax algorithm as follows:

$$\delta^o = \min_K \{\max_i (\delta_i)\} \rightarrow K^o \text{ and } b^o.$$

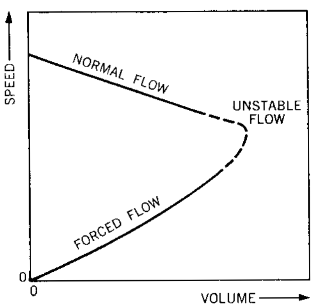
Traffic volume and through-bands

volume

As indicated earlier, a highest-speed and lowest-cycle combination is selected within allowable limits for both speeds and cycle times. A further restriction for both algorithms described can be imposed by considering the speed/volume relationships.

When considering traffic flow on a freeway or in a tunnel, we observe that, at relatively high speeds, wider vehicular spacings reduce the traffic density. Empirical and theoretical studies[2] of such traffic-flow characteristics show that traffic volume (i.e., vehicles per hour) decreases with increasing speeds. This relationship holds true for normal free flow. At a point of maximum volume at a reduced speed, flow becomes unstable and finally turns into a forced flow. In this region, speed and volume decrease simultaneously. Figure 5 shows a typical speed/volume curve. The program accepts only a single-valued function of volume, so that only one section, either the normal or forced flow part, can be considered.

Figure 5 A speed/volume curve



Such a speed/volume relation can be established for any urban artery as if all signalized intersections were assumed to be green and continuous free flow permitted for an extended period. These curves represent the road (or lane) flow capacity as a function of speed only. They may differ considerably in values represented, depending on road conditions (e.g., slipperiness, visibility). The volumes obtained from traffic counts on signalized arteries are obviously not for free flow. They are usually given for pulsed flow, or platoon flow, which results from signal *flow gating*. The following gating relationship can be assumed for an intersection:

$$\frac{\text{Actual (pulsed) volume}}{\text{Free (capacity) volume}} \cdot F = \text{green ratio} \qquad F \geq 1,$$

where F is a safety or inefficiency factor, which may be used to compensate for acceleration and deceleration losses due to signalization and also for inaccuracy of the assumed speed/volume curve. F may also roughly compensate for the errors due to neglect of intersection widths and related crossing times. Although it is possible to develop theoretical expressions to account for these effects, it is felt that, in practice, it suffices to relate F inversely to cycle length. The speed/volume curve is employed to estimate free flow volumes for given speeds.

Rewriting the above gating relationship, we have

$$\frac{\mathcal{A}_{iL}}{\mathcal{F}_{iL}} \cdot F = \mathcal{G}_{iL} = \frac{G_{iL}}{c},$$

where, again,

$$\mathcal{F}_L = f(V_L) = \mathcal{F}_L(V_L),$$

as defined by the speed/volume curve.

Now let $\mathcal{A}_{iL}$ be the required volume (flow demand) approaching intersection I in direction L. Then the required free flow volume is

$$\mathcal{F}_{iL} = \frac{\mathcal{A}_{iL}}{\mathcal{G}_{iL}} \cdot F. \tag{9}$$

The required equivalent free-flow volume for the artery is determined by the bottleneck intersection, which is the one with a green ratio $\mathcal{G}_{mL}$ and a demand $\mathcal{A}_{mL}$, which yield the largest free flow volume:

$$\mathcal{F}_L = \frac{\mathcal{A}_{mL}}{\mathcal{G}_{mL}} \cdot F. \tag{10}$$

Checking against the speed/volume curve, we find the maximum speed (if any) at which such equivalent free-flow volume is feasible. If no such volume is feasible, it may be possible to increase the green ratios sufficiently to allow the volume demand at a reasonable speed. This may impose impractical restriction of cross-street traffic flow. In such cases, the arterial flow demand cannot be met and some compromise is required. In a properly controlled system, it may be desired to allow demands (entering traffic) only to the extent that normal flow (at above minimum speeds prior to flow breakdown) can be maintained.

Figure 6 shows a typical mid-green relationship for such transposed intersections. Again we transpose the base midpoint along both bands to the intersection under consideration, so that

$$P = c + \tfrac{1}{2}G_b$$

yields

$$P - t_{i2} = c + \tfrac{1}{2}G_b - t_{i2}$$

and

$$P + t_{i1} = c + \tfrac{1}{2}G_b + t_{i1},$$

where

$$t_{i1} = \frac{D_i}{V_1}, \qquad t_{i2} = \frac{D_i}{V_2},$$

and

$$T_i = t_{i1} + t_{i2}.$$

These two transposed points represent the band midpoints and as such may be slightly shifted as the band is being narrowed by red-interval interferences. The intersection's mid-green points are moved as close as possible to both band midpoints. Offset positioning considers the unavoidable deviations Δ_{i1} and Δ_{i2}. The designer may select equal deviations or assign values inversely related to traffic volumes. Generally, we assume some weighting factor α_i and $(1 - \alpha_i)$ for direction 1 and 2, respectively, where $0 < \alpha_i < 1$.

Thus, we require:

$$\alpha_i \Delta_{i1} = (1 - \alpha_i)\Delta_{i2}. \tag{11}$$

But

$$\begin{aligned}\Delta_{i1} &= \phi_i + c + \tfrac{1}{2}G_i - (P + t_{i1}) \\ &= \phi_i + \tfrac{1}{2}(G_i - G_b) - t_{i1}, \\ \Delta_{i2} &= P - t_{i2} - (\phi_i + \tfrac{1}{2}G_i) \\ &= c - t_{i2} - \tfrac{1}{2}(G_i - G_b) - \phi_i,\end{aligned}$$

where ϕ_i is the offset for intersection I.

Substituting in Equation 11 and solving for ϕ_i, we obtain

$$\phi_i = t_{i1}\alpha_i + (c - t_{i2})(1 - \alpha_i) - \tfrac{1}{2}(G_i - G_b).$$

This formula provides a *least deviations* positioning for all intersections that transpose into the c zone (Figure 6), and it yields a *lead-trail* interference configuration. This c zone is defined by

$$\frac{c}{2} \leq T_i < c.$$

For intersections that transpose into the $c/2$ zone, where $0 \leq T_i \leq c/2$, a similar development yields

$$\phi_i = t_{i1}\alpha_i + (c - t_{i2})(1 - \alpha_i) - \tfrac{1}{2}(G_i - G_b) + c\alpha_i$$

which provides for a *trail-lead* configuration.

When extreme values of α (close to zero or one) are used, it is possible to interfere with one band while the other is given excess green time (outside of band). Such positioning is not permitted and the algorithm overrides the design required by α to take full advantage of the available green phase.

Extensions

curvilinear fit

The following extensions to the algorithms described may yield significant results. These extensions may be incorporated in future versions of the program.

A curvilinear fit for a set of system constants or speeds along an artery may provide better results. Such a fit implies that, instead of the fixed speed along an arterial direction, speeds varying from location to location would be imposed. In this case, two alternative approaches are considered:

1. Obtain a second (or higher) order curvilinear fit, then approximate this fit by connecting line segments, perferably improving on the curvilinear fit.
2. Develop a simple linear fit for the first three intersections, and test for goodness of fit; if good, include the fourth intersection, and test again; if not good, the first segment is composed of the first two intersections only. Continue until an appropriate number of segments for all intersections is developed.

In both cases, the smallest allowed cycle time that meets the volume demand (via the speed/volume relation) is picked. For each segment and its systems constant K_j, a speed is then calculated for the condition

$$K_j = \frac{c}{2} V_j.$$

The actual speeds can be selected in relation to volume demands in both directions. The offset design does not change, with the exception that each segment has its own base. Offsets can be easily transformed to a common reference time at, say, the first intersection. As long as each segment meets its volume requirements, it may be desirable to obtain different bandwidths for different segments.

network control

The previous modification leads to a further extension, i.e., the adaptation of such a method to network control.

Consider a network element of 2 × 2 intersections. Although we can design a progression around three legs of such an element, we are normally faced with the problem of *nonclosure* if only one speed is considered. If we allow sufficient (but practical) speed change around the element, closure can be obtained on some best-fit basis.

The extension to larger networks does not pose a conceptual problem, although it may become quite complex computationally. Such a time/space approach to the network control problem may not yield as good results as with an "instantaneous decision" control system, but it could require considerably less equipment.

Computer program

A computer program[3] was written in FORTRAN-language for the IBM 1620, 7040, and 7090 systems. The input data accepted and utilized by the program include:

- Speed/volume curve values
- Intersection distances
- Number of effective lanes
- Cycle time range
- Desired cycle time precision
- Volumes approaching each intersection
- Minimum pedestrian crossing times
- Various program option controls, such as left-turn phase selection, double cycle control, etc.

Initially, green ratios are developed from volumes and minimum pedestrian crossing times. Then a set of alternative system constants are developed. Next, minimum cycle time and maximum speed are computed within the restrictions of given volume demands and speed/volume relation. Finally, offsets that provide through-bands with maximum volumes are computed for all intersections.

Output is provided in terms of selected system cycle time, average speeds in both arterial directions, feasible system volumes, and intersection green splits and offsets. Time-space diagrams can be machine-plotted if an IBM 1627 plotter is available.

The complete solution (with card-punched results) for a 15- to 20-intersection arterial progression requires 20 to 30 minutes on a 1620 MODEL I (40K and 1311 disk file) system. Larger problems can be handled with an IBM 7000 series system in a few minutes.

CITED REFERENCE AND FOOTNOTES

1. A similar version of this paper was presented at the Annual Meeting of the Institute of Traffic Engineers, Miami Beach, Florida, November 1964.
2. R. T. Underwood, "Some aspects of the theory of traffic flow," *Australian Road Research*, Ramsay, Ware Publishing Pty., Ltd., North Melbourne, N. I., No. 2, 35–47 (June 1962).
3. Two versions of the program, called "Vehicular Traffic Control Time/Space Diagram (Progression) Design Program," can be ordered through any IBM branch Office. The 1620 version, PID 9.2.065, is included in the IBM 1620 General Program Library. The 7040/7090 version, PID 7040 T4IBM0005, is part of the SHARE General Program Library.