REFERENCES

1. W.C. Jordan and M.A. Turnquist. Zone Scheduling of Bus Routes to Improve Service Reliability. Transportation Science, Vol. 13, No. 3, 1979, pp. 242-268.
2. M.D. Abkowitz. The Impact of Service Reliability on Work Travel Behavior. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA, Ph.D. dissertation, 1980.
3. J.N. Prashka. Development of a "Reliability of Travel Modes" Variable for Mode-Choice Models. Transportation Center, Northwestern Univ., Evanston, IL, Ph.D. dissertation, 1977.
4. R.E. Allsop. Delay at a Fixed Time Traffic Signal--I: Theoretical Analysis. Transportation Science, Vol. 6, 1972, pp. 260-285.
5. D.L. Gerlough and M.J. Huber. Traffic Flow Theory. TRB, Special Rept. 165, 1975.
6. G.F. Newell. Approximation Methods for Queues with Application to the Fixed-Cycle Traffic Light. SIAM Review, Vol. 7, 1965, pp. 223-237.
7. T.E.H. Williams and J. Emmerson. Traffic Volumes, Journey Times and Speeds by Moving-Observer Method. Traffic Engineering and Control, Vol. 3, 1961, pp. 159-162 and 168.
8. A.J. Miller. Settings for Fixed-Cycle Traffic Signals. Proc., Australian Road Research Board Conference, Vol. 2, No. 1, 1964, pp. 342-362.
9. G.E.P. Box and G.M. Jenkins. Time Series Analysis: Forecasting and Control. Holden-Day, New York, 1970.
10. N.W. Polhemus. On the Statistical Analysis of Non-Homogenous Traffic Streams. Transportation Planning and Technology, Vol. 6, 1980, pp. 45-54.
11. D.L. Gerlough and F.C. Barnes. Poisson and Other Distributions in Traffic. Eno Foundation for Transportation, Saugatuck, CT, 1971.
12. S.G. Ritchie. Bus and Car Pool Lanes on Arterial Roads--A Simulation Study. Proc., Transportation Research Forum, 21st Annual Meeting, Philadelphia, Oct. 1980, Richard B. Cross Co., Oxford, IN, 1980, pp. 280-287.
13. D.J. Buckley. Road Traffic Counting Distributions. Transportation Research, Vol. 1, 1967, pp. 105-116.
14. E. Williamson and M.H. Bretherton. Tables of the Negative Binomial Probability Distribution. Wiley, London, 1963.
15. J.R. Benjamin and C.A. Cornell. Probability, Statistics, and Decision for Civil Engineers. McGraw-Hill, New York, 1970.
16. E. Kreyszig. Advanced Engineering Mathematics, 4th ed. Wiley, New York, 1979.
17. P.G. Hoel. Introduction to Mathematical Statistics. Wiley, New York, 1966.
18. A. Karlquist and B. Marsjo. Statistical Urban Models. Environment and Planning, Vol. 3, 1971, pp. 83-98.
19. F.J. Anscombe. Sampling Theory of the Negative Binomial and Logarithmic Series Distributions. Biometrika, Vol. 37, 1950.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

Delay Models of Traffic-Actuated Signal Controls

FENG-BOR LIN AND FARROKH MAZDEYASNA

Traffic-actuated signal controls have more control variables for engineers to deal with than a pretimed control. The increased sophistication in their control logic provides greater flexibilities in signal control but also makes the evaluation of their performance more difficult. At the heart of the problem is that traffic delays cannot be readily related to the control variables of a traffic-actuated control. This prompts practicing engineers to rely mostly on short-term, subjective field observations for evaluation purposes. To provide an improved capability for evaluating alternative timing settings, delay models are developed in this study for semiactuated and full-actuated controls that employ motion detectors and sequential phasing. These models are based on a modified version of Webster's formula. The modifications include the use of average cycle length, average green duration, and two coefficients of sensitivity that reflect the degree of sensitivity of delay to a given combination of traffic and control conditions. Average cycle length and average green duration are dependent on the settings of the control variables and the flow pattern at an intersection. They can be estimated by existing methods.

Traffic-actuated controls employ relatively complex logic to regulate traffic flows. This type of logic infuses a much-needed flexibility into signal control, but it also makes the performance evaluation of a traffic-actuated control difficult. A major problem is that traffic delays resulting from such a control cannot be readily related to the settings of the control variables and the flow pattern at an intersection.

Current understanding of traffic delays at a traffic-actuated signal is obtained only through sensitivity analyses with the aid of computer simulation models. Tarnoff and Parsonson (1) have provided a detailed review of the findings of these simulation studies. Computer simulation models, however, have significant limitations.

For one thing, practicing engineers may not be familiar with the nature and the capability of such models. Furthermore, to ensure broad applicabilities, simulation models are often difficult to use in terms of data needs, requirements of computer facilities, and the time one has to spend to learn how to use them. As a result, practicing engineers still rely mostly on short-term, subjective field observations in evaluating timing settings.

An alternative to the use of computer simulation is to develop a model in the form of a formula or a set of formulas. Such a model would allow expedient evaluation of a large number of alternatives and would be particularly useful in searching for optimal ways of using a signal control. This in turn could encourage practicing engineers to improve the efficiency of existing signal controls.

To partly satisfy this need, this paper presents a set of delay models for semiactuated controls and full-actuated controls that employ motion detectors. These delay models are calibrated in terms of simulation data. They are applicable to signal controls at individual intersections when single-ring, sequential phasing is used. The traffic flows considered in this study include only straight-through passenger vehicles. Other types of vehicles can be transformed into equivalent straight-through passenger vehicles for analysis (2).

GENERAL FORMULATION

The delay models assume the following form:

$$D = 0.9 \{[C(1 - Ax)^2/2(1 - AxBy)] + [3600 (By)^2/2Q(1 - By)]\} \quad (1)$$

where

- D = average delay (s/vehicle),
- C = average cycle length (s),
- x = ratio of effective green to average cycle length = G_e/C,
- G_e = effective green = $G + Y - k$ (s),
- G = average green duration (s),
- Y = yellow duration (s),
- k = loss time per phase,
- Q = traffic volume in a lane (vph),
- y = saturation ratio = $QC/(Q_sG_e)$, and
- Q_s = saturation flow rate, which is approximately 1800 vph of effective green.

Equation 1 is a modified version of Webster's formula (3). Its unique feature is the inclusion of the two coefficients of sensitivity, A and B. The reason for including these coefficients is that a vehicle subjected to a traffic-actuated control can exert an influence in the transfer of right-of-way. Consequently, traffic-actuated delays and pretimed delays can be expected to have different sensitivities to both the ratio of effective green to cycle length and the saturation ratio. A larger value of A represents a lesser degree of sensitivity of the average delay to the ratio of effective green to cycle length (x). In contrast, a larger value of B implies a higher degree of sensitivity of the average delay to the saturation ratio (y).

SIMULATION MODEL

The simulation model used in this study comprises a flow processor and a signal processor. The flow processor generates vehicle arrivals and processes the vehicles through an intersection according to the signal indications. The speed, location, and acceleration rate of each vehicle are updated by this processor once every second. The operation of this processor is a function of the signal indications. It is independent of the type of signal control. Based on the control logic of a given type of signal control, the signal processor uses flow data provided by the flow processor to determine the signal indications for each 1-s scanning interval.

Under various combinations of flow and pretimed signal settings, average delays obtained from the simulation model are generally within 15 percent of the values estimated from Webster's formula. This indicates that the flow processor can move vehicles downstream in a reasonably reliable manner. The average greens of individual phases of semiactuated controls and full-actuated controls as generated from the signal processor are found to be within 2-3 s of estimates obtained from analytical models (4,5). In addition, the simulated delays for a full-actuated control under near-optimal control conditions agree very well with estimates obtained by Morris and Pak-Poy (6). These tests do not constitute a complete validation of the simulation model. Nevertheless, they confirm the ability of the simulation model to produce satisfactory data.

DELAY MODELS

Semiactuated controls may be used when a lightly traveled side street intersects a major street. Detectors are installed on the side street to collect flow data for making signal-timing decisions. When motion detectors are used, the key control variables of this type of control usually include minimum green (G_{min}) for the major street and initial portion (I), unit extension (U), maximum allowable green (G_{max}), and detector setback (S) for the side street. Without vehicle actuation of the detectors, the green light is always given to the major street.

In contrast, full-actuated controls require the use of detectors on all approaches that are subjected to signal control. When motion detectors are used, the duration of each green phase in a given cycle is governed by the same set of control variables, which usually includes I, U, G_{max}, and S as defined previously.

To calibrate Equation 1 for either type of the control, three levels of the initial portion were considered: 5, 8.5, and 12.5 s. At each level more than 90 different combinations of flow conditions and settings of the control variables were examined through computer simulation. To avoid unnecessary complications, only two-phase controls were dealt with. Each phase involves two traffic lanes with equal or unequal volumes. The unit extension is confined to a value between 3 and 6 s, and the detector setback is from 50 to 120 ft.

Under semiactuated controls, the minimum green for the major street varies from 20 to 50 s and the maximum green for the side street is limited to 30 s.

For each combination of the flow and signal control conditions, the simulation model generates average delays, average cycle length, and average green duration related to each signal phase. The generated data were used in Equation 1 to determine the combination of A and B that best duplicates the simulated delays.

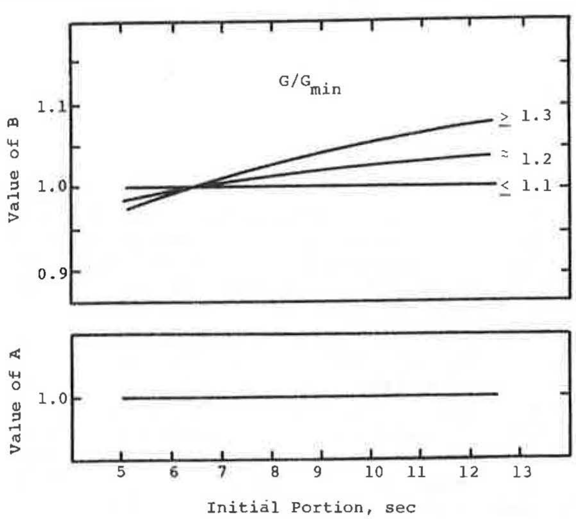
For semiactuated controls, the data generated for the side-street traffic were analyzed separately from those for the major-street traffic. The reason for this is that semiactuated controls are only responsive to the side-street vehicles and thus the side-street delays and the major-street delays are likely to be affected by the controls in different ways.

The results of the model calibration are shown in Figures 1 and 2 for semiactuated controls and in Figure 3 for full-actuated controls. Figure 1 shows that for semiactuated control, both A and B for the side-street delays are greater than or equal to 1.0. This indicates that the side-street delays are less sensitive to the ratio of effective green to cycle length (x) and more sensitive to the saturation ratio (y) than pretimed delays. Since B decreases with respect to the initial portion, it can also be said that a shorter initial portion gives rise to a greater sensitivity of the delays to the saturation ratio. In other words, when a short initial portion is used, the average side-street delay increases rapidly with respect to the saturation ratio. When the G/G_{max} ratio increases, however, A and B approach 1.0. This implies a convergence of semiactuated controls to pretimed controls.

For the major-street traffic, A has a constant value of 1.0 (Figure 2) as in the case of a pretimed control. The value of B increases with the initial portion used for the side-street traffic. Therefore, a longer initial portion for the side street could cause the major-street delays to rise rapidly with the saturation ratio. The value of B for the major-street traffic also varies with the ratio of average

Figure 1. Calibrated values of A and B for side-street traffic under semiactuated control.

Figure 2. Calibrated values of A and B for major-street traffic under semiactuated control.

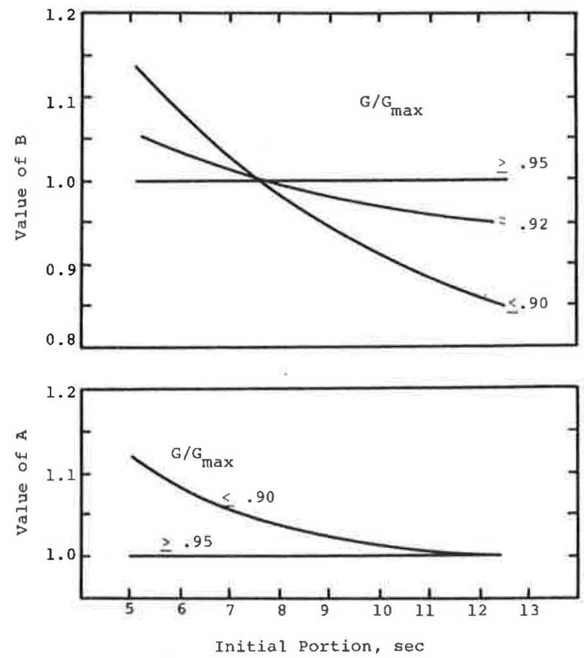


green to minimum green (G/G_{min}). When this ratio is less than 1.1, the variation of the major-street green from one cycle to another is small and B approaches 1.0.

Under full-actuated controls, the value of A is at least as large as 1.0, and it can be concluded that average delays resulting from a full-actuated control are less sensitive to the G_e/C ratio than pretimed delays.

The value of B exceeds 1.0 when the initial portion is below 7.5 s and drops under 1.0 at a longer initial portion. This feature signifies that the

Figure 3. Calibrated values of A and B for full-actuated control.



sensitivity of the average full-actuated delays to the saturation ratio increases when the initial portion decreases. It follows logically that a heavier flow should be given a longer initial portion. Also, the average full-actuated delays are not so adversely affected by the saturation ratio as pretimed delays when the initial portion is greater than 7.5 s. The advantage of full-actuated controls, however, disappears when the G/G_{max} ratio is in excess of approximately 0.95. At this level of the ratio, full-actuated controls behave more or less like pretimed controls.

To be useful as a tool for evaluating alternative timing settings and detector setbacks, Equation 1 has to be used in conjunction with a method for estimating G and C. A reasonable option is to use the methods presented by Lin (4,5) in two recent studies. These methods relate the average green of a signal phase of a semiactuated control or a full-actuated control to the control variables and the traffic flow pattern at an intersection. The formulations of the methods are not simple because of the complex logic of the traffic-actuated controls. Nevertheless, the methods can be applied manually or be implemented in the form of short computer programs with about 60 FORTRAN statements for semiactuated controls and about 100 statements for full-actuated controls.

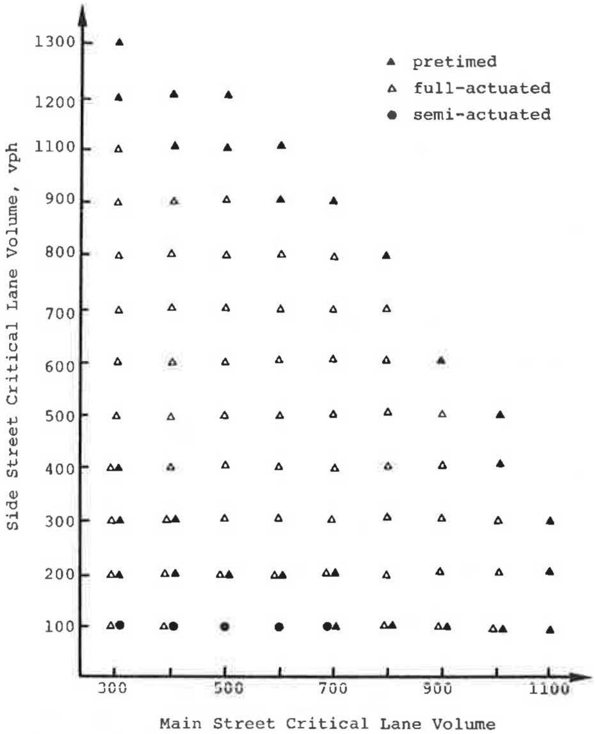
It should be noted that, in their existing forms, these estimation methods are generally applicable only when a unit extension of greater than approximately 3-3.5 s is used. With a shorter unit extension, premature termination of green phases is quite likely and the methods will require modifications.

Based on these methods for estimating G and C, the average delays obtained from the delay models agree reasonably well with the values generated from the simulation model. The differences are within 3 s in more than 85 percent of the cases examined in this study.

APPLICATIONS

The primary applications of the delay models are in

Figure 4. Domains of preferred types of two-phase signal control derived from delay models and Equation 1.



the evaluation of alternative timing settings. By using the models as a tool for sensitivity analysis, one can examine relatively easily how changes in the timing settings and detector setbacks may affect the efficiency of a signal control. The models are particularly useful for searching for optimal controls when a microcomputer is available to implement the methods for estimating G and C.

The delay models can also be used to assist in the selection of alternative types of signal control. For example, by using the delay models and Webster's formula, one can obtain Figure 4. This figure shows the most efficient types of signal control along pretimed, semiactuated, and full-actuated controls for various combinations of flows.

CONCLUSIONS

Use of simulation models for evaluating a large number of alternative timing settings and detector setbacks is usually cumbersome and requires substantial resources. The delay models described in this paper provide a more efficient alternative. These models, however, are applicable only when motion detectors are used and when sequential phasing is in effect. Similar models may also be developed for other types of traffic-actuated control. The availability of such simplified models could encourage practicing engineers to make an effort to improve existing controls.

Discussion

Kenneth G. Courage

Within the scope stated by the authors, the study appears to be sound. The methodology is scientific and the results are reasonable.

The authors propose an extension to Webster's delay model to deal explicitly with certain operating parameters (minimum and maximum green) for a traffic-actuated controller. They suggest that a model of this form is preferable to existing models that treat these parameters implicitly. The results of the proposed model are not compared with those of the existing models. Such a comparison would have made the results more credible. Greater credibility might also have been achieved by starting with the TRANSYT modification to Webster's model, which deals with oversaturated as well as undersaturated operation.

The applicability of these results is constrained by the scope of the study, which was limited to exclude volume density operation and presence detection on the approaches. These features are both very common and both have been shown to produce a more efficient signal operation than the conventional actuated controller with motion detection.

Another limiting factor in the results is that the coefficients A and B are shown to vary with the parameter G/G_{max}. In other words, the optimal setting of the operating parameters varies with traffic volume; therefore no permanent controller settings can be developed by using the proposed model. This variation has been recognized in the past and was the primary motivation behind the development of the volume density controller.

It must be recognized that the most successful modes of locally actuated intersection control are based on intuitive mechanical models. These models are primitive and they defy purely analytical treatment. Their popularity is derived from the fact that they can be fully implemented on the street, whereas theoretical models, such as the one discussed in this paper, cannot.

Authors' Closure

The operation of the pulse-mode traffic-actuated control has not been properly modeled and analyzed in the past. A comprehensive discussion of this issue is not appropriate for this closure. Nevertheless, an existing model discussed by Courage and Papapanou (7) can be used for a short discussion. This earlier model is based on a control strategy that has the following characteristics: (a) it distributes available green time in proportion to demand on critical approaches, and (b) it minimizes wasted time by terminating each green interval as soon as the queue of vehicles has been properly serviced. We indicated that this control strategy closely approximates the operation of the traditional traffic-actuated controller that has been properly timed and that the delay estimates will therefore reflect the best operations that can be expected from traffic-actuated control. The validity of these claims aside, this earlier model is aimed at estimating the minimum delays. In contrast, the proposed model provides a mechanism for estimating delays under various combinations of traffic and control conditions.

Furthermore, a close examination of the earlier model will reveal that the only control variable considered in the model is the maximum green of a signal phase. Unfortunately, this variable is of concern only under very limited conditions. The model also uses average cycle length (C_a), which is calculated as $C_a = L/(1.0 - Y)$, where L is the total loss time per cycle and Y is the overall degree of saturation of critical movements. Again, under a traffic-actuated control, the average cycle length cannot be adequately estimated from such an equation. The result is a model that has little to do with the actual timing settings and detector setback.

In short, the proposed and the earlier models have distinct characteristics. A comparison of the two models is really meaningless and will not make the proposed model either more or less credible.

The discussant suggested that the simulation model used to develop the delay models should have been calibrated on the basis of the TRANSYT-7F model (8) rather than on the basis of Webster's delay model (3). The contention was that Webster's model gives unrealistically high estimates of delay when the saturation ratio approaches or exceeds 1.0 (Figure 5). Such a contention reflects a general lack of understanding, not only about the nature of Webster's model, but also about the flow characteristics at saturation ratios near or exceeding 1.0.

Figure 5. Comparison of Webster's model with TRANSYT-7F model.

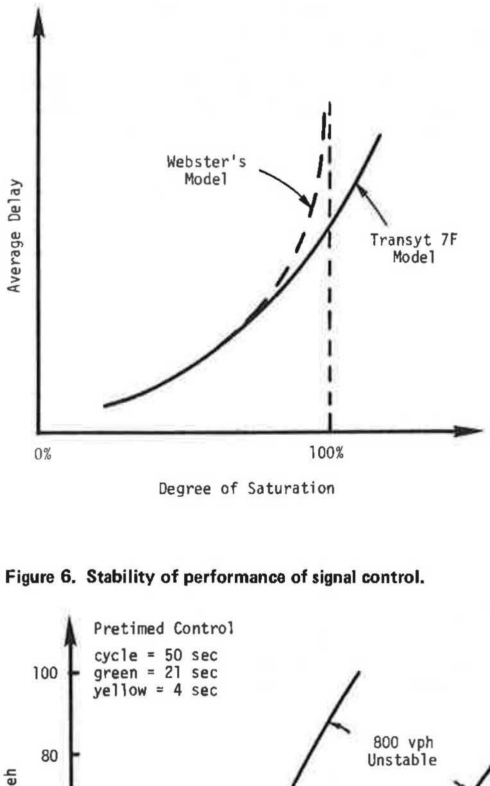
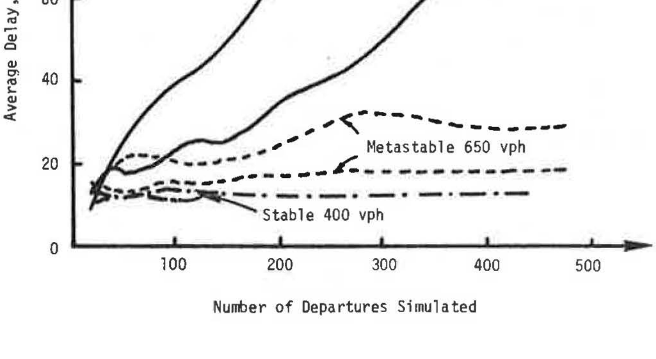


Figure 6. Stability of performance of signal control.



To resolve this issue, it is necessary to point out that in terms of delays, the operation of a signal system can be classified into the following states:

1. Stable state: In this state the average delay of a flow is primarily a function of the flow rate. Variations in the delays from one field observation or simulation run to another are small as long as the flow rate and the control conditions remain unchanged.
2. Metastable state: In this state the average delay depends not only on the flow rate but also on the sequence of the arriving headways. If the same flow rate persists, a stable value of the average delay can still be obtained. However, the variations in the delays become substantial at a given flow rate.
3. Unstable state: In this state the average delay depends not only on the flow rate and the sequence of the arriving headways but also on the time period in which a given flow rate persists. In other words, the average delay is time dependent. The longer the flow rate persists, the higher the average delay becomes.

The existence of these states can be identified from computer simulation. Figure 6 shows an example. There are no clear-cut boundaries between the various states. For pretimed control, the metastable state may arise when the saturation ratio is in the range of 0.8-0.9; the unstable state may emerge when the ratio is about 0.9 or greater.

Webster's model gives the estimated delay for a flow that persists indefinitely. Naturally, when the saturation ratio approaches 1.0, the estimated delay approaches infinity. If this nature of the model is not recognized, the comparison between Webster's model and the TRANSYT-7F model is just like the comparison between apples and oranges.

Since in the real world a flow will never persist long enough to induce an infinite average delay, the TRANSYT-7F model attempts to account for this fact by using a delay function with finite delay values (Figure 5). In so doing, it only gives one a false sense of security. The reason is that at high saturation flow rates, delay is time dependent. Therefore, in a finite time frame, the single delay function of the TRANSYT-7F model should be replaced with a set of delay functions, each of which is associated with a given time period of signal operation. In reality, this is extremely difficult if not impossible to accomplish.

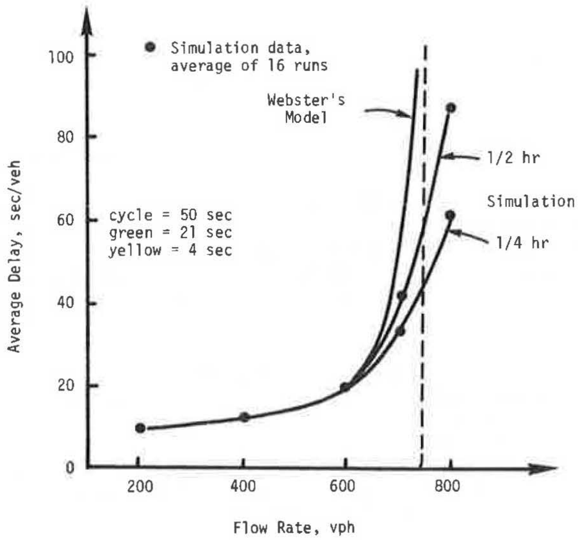
In the absence of better information, Webster's model is a reasonable basis for calibrating a simulation model. The calibration, of course, should be based on stable or metastable operations. A simulation model calibrated in this manner will not result in unrealistically high average delays over a finite time period. Figure 7 illustrates this feature.

In summary, the TRANSYT-7F model does not have real advantages over Webster's model. Significant improvements can be made only if the delay function at high saturation ratios can be explicitly and reliably related to time.

We recognized that the proposed models are not applicable to volume-density operation and presence-mode operation. However, the same methodology can be employed to develop a model for either one of such operations.

The discussant seemed to object to the fact that the model showed that the optimal settings of the operating parameters varied with traffic volume. Such variations not only are inherent to pulse-mode operation but also exist in presence-mode operation and volume-density operation. In fact, as long as a signal control requires predetermined settings, it

Figure 7. Comparison of simulated delays with Webster's delay under pretimed control.



will not always result in an optimal control under varying traffic-flow conditions. Precisely for this reason, it is desirable to have a reasonably simple yet reliable model to determine the trade-off of timing settings with respect to a typical daily flow pattern. Such a trade-off analysis would enable one to select permanent settings. The use of volume-density control alleviates but does not eliminate this problem.

It would certainly be a blessing if the operation of a signal control could be adequately represented by a primitive and intuitive model. In reality, such intuitive models for analyzing traffic-actuated controls are often misleading and are usually not better than practicing engineers' intuitive judgments. In anticipation of increased use of microcomputers, there is room for developing models that are more reliable than intuitive models but less difficult to use than most existing simulation models.

REFERENCES

1. P.J. Tarnoff and P.S. Parsonson. Selecting Traffic Signal Control at Individual Intersections. NCHRP, Rept. 233, 1981.
2. Interim Materials on Highway Capacity. TRB, Transportation Research Circular 212, Jan. 1980.
3. F.V. Webster. Traffic Signal Settings. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, TRRL Rept. 39, 1961.
4. F.B. Lin. Predictive Models of Traffic-Actuated Cycle Splits. Transportation Research, Vol. 16, Part B, No. 5, 1982, pp. 361-372.
5. F.B. Lin. Estimation of Average Phase Duration of Full-Actuated Signals. TRB, Transportation Research Record 881, 1982, pp. 65-72.
6. R.W.T. Morris and P.G. Pak-Poy. Intersection Control by Vehicle-Actuated Signals. Traffic Engineering and Control, Oct. 1967, pp. 288-293.
7. K.G. Courage and P.P. Papapanou. Estimation of Delay at Traffic-Actuated Signals. TRB, Transportation Research Record 630, 1977, pp. 17-21.
8. TRANSYT-7F User's Manual. FHWA, Dec. 1981.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

Another Look at Bandwidth Maximization

KARSTEN G. BAASS

One solution to the problem of fixed-time traffic signal coordination is the provision of a large green band that allows road users to drive at a reasonable speed without stopping. This solution is popular with drivers, although it does not necessarily lead to delay minimization except in special cases. A method for deriving the globally maximal bandwidth together with all possible suboptimal values is described. The programs WAVE1 and WAVE2 can also be used to generate curves that show the continuous relation between uniform progression speed and corresponding maximal bandwidth over a wide range of speeds and cycles. The typical shape of this bandwidth-speed relationship is explained theoretically, and the theory is used in the development of the algorithm. It is shown that bandwidth varies greatly with progression speed and it is suggested that setting bandwidth at the globally optimal value may not always be the best choice. The decision to adopt a progression speed, a bandwidth, and a cycle time should take into account a range of values of speed and cycle. The proposed method was applied to 18 data sets of up to 24 intersections taken from the published literature and the results obtained were compared with those given by the mixed-integer linear-programming approach. Computer execution time is extremely short and the storage space required is negligible, so the method could be of interest in practical applications.

The maximization of bandwidth is one of the two approaches used for determining offsets between fixed-time traffic lights on an artery. There are a number of fairly restrictive hypotheses related to this approach, e.g., the assumptions of a uniform platoon, no platoon dispersion, low volumes, and no or very few cars entering the artery from side streets. Situations corresponding to these assumptions are rare. Nevertheless, the bandwidth-maximizing approach is psychologically attractive to the user, who is unable to distinguish between a nonsynchronized artery and one that is perfectly synchronized for delay and stop minimization but does not allow the user to pass at a reasonable speed through the artery without stopping.

Little and others (1) and Morgan (2) were the first to suggest a mathematical formulation for the bandwidth-maximizing problem, and more recently Little and others (3) published a program called MAXBAND. This program is based on a mixed-integer linear-programming approach and determines the speeds that give the overall maximum bandwidth over a range of acceptable speeds. The linear-programming approach also allows for variations in speed between intersections and enables new constraints to be easily introduced.

This paper describes an algorithm that determines the overall maximum bandwidth together with all suboptimal values, if they exist, for a wide range of speeds. At this time, only two-phase fixed-time