RESEARCH REPORT NO. 46

Traffic Signal Synchronization for High Flows on a Two-Way Street

G. F. Newell

THE INSTITUTE OF TRANSPORTATION
AND TRAFFIC ENGINEERING
UNIVERSITY OF CALIFORNIA

F. C. BARNES

Institute of Transportation
and Traffic Engineering

Research Report No. 46

TRAFFIC SIGNAL SYNCHRONIZATION FOR HIGH FLOWS ON A TWO-WAY STREET

G. F. Newell

This work was supported in part by
the National Science Foundation
under grant GP-6459

University of California
Berkeley, August 1967

ABSTRACT

Suppose that signals on a two-way main street are constrained to operate with a common cycle time and equal splits. We wish to choose the offsets so as to minimize both the total delay and the total number of stops to the main-street traffic. The problem of estimating these offsets is, in general, quite complicated, but if the flows in both directions are at saturation, the optimal offset between a pair of adjacent signals does not depend upon the offsets between other pairs of signals. The mathematical problem reduces to an analysis of delays and stops between only a single pair of intersections.

Delays are evaluated from a fluid model with no platoon spreading. For unequal flows in the two directions, the optimal offsets are those which give a progression to the direction of heavier traffic. If the flows are equal, however, there is a wide range of setting which give the same delays and stops. For closely spaced intersections the choice of zero offsets is included among the optimal settings. It has the additional advantage that it gives equal delays to the two directions.

INTRODUCTION

To determine for a network of roads the signal settings which minimize delays or stops is a very complex problem. To gain some insight into the qualitative aspects of this problem, it is therefore helpful to investigate some simple idealized examples. In a previous paper, Baverez and the present author[1] considered some aspects of synchronization of signals on a one-way street. Here we consider some features of the more complex problem for a two-way street at flows close to saturation, and with no turning traffic.

In practice traffic engineers usually set signals on a two-way street according to one of the following schemes: (a) set all signals to operate on the same phase (simultaneous greens and reds), (b) set signals with alternating settings, each signal has a half-cycle offset from its neighbors, (c) give a maximum through-band for the direction of heavier traffic flow and, if there is any option left, do the best one can for the other direction, or (d) try to produce non-zero bandwidths in both directions and perhaps maximize the sum of the two bandwidths.

The last of these schemes has received considerable attention recently.[2,3,4] There are now a variety of computer algorithms for determination of settings which give a maximum total bandwidth. Before computers there were also many graphical or trial-and-error methods for finding maximal or near maximal bands. Despite the popularity of this problem, however, it does not obviously lead to desirable settings. The capacity of a signalized street is determined mainly by the green times, particularly the shortest green time, and is almost always larger than the capacity of the through-band. In many cases the bandwidth is only a small fraction of the green time. It will, therefore, not be uncommon to have a flow which exceeds the capacity of the through-band but does not exceed the capacity of the highways. In such cases, the system is capable of carrying this steady flow with only finite delays, but no driver can take advantage of the through-band. It is certainly not obvious that the maximization of the bandwidth is equivalent to minimization of delays or stops or any other objective with some economic significance.

In order to estimate delays, stops, etc. on a two-way street we will adopt here the commonly used (but not necessarily realistic) postulate that all cars traveling between the same pair of adjacent intersections (in the same direction) have the same travel time. Equivalently, they all have the same average velocity so there is no spreading of a platoon. We also treat the traffic as if it were a continuous fluid which has a maximum flow rate of s during discharge from a queue, the same at every intersection (more generally we may have s in one direction and $s' \neq s$ in the other direction).

The standard way of representing the flow of traffic graphically is to draw space-time trajectories of the cars as in Fig. 1. We usually draw these as if cars travel at a constant velocity between any pair of intersections (not necessarily the same for every pair, however). But actually it is only the travel times between intersections that enter into the analysis, so the detailed shape of the trajectories between intersections is irrelevant. This is merely a convenient way of picturing the travel times.

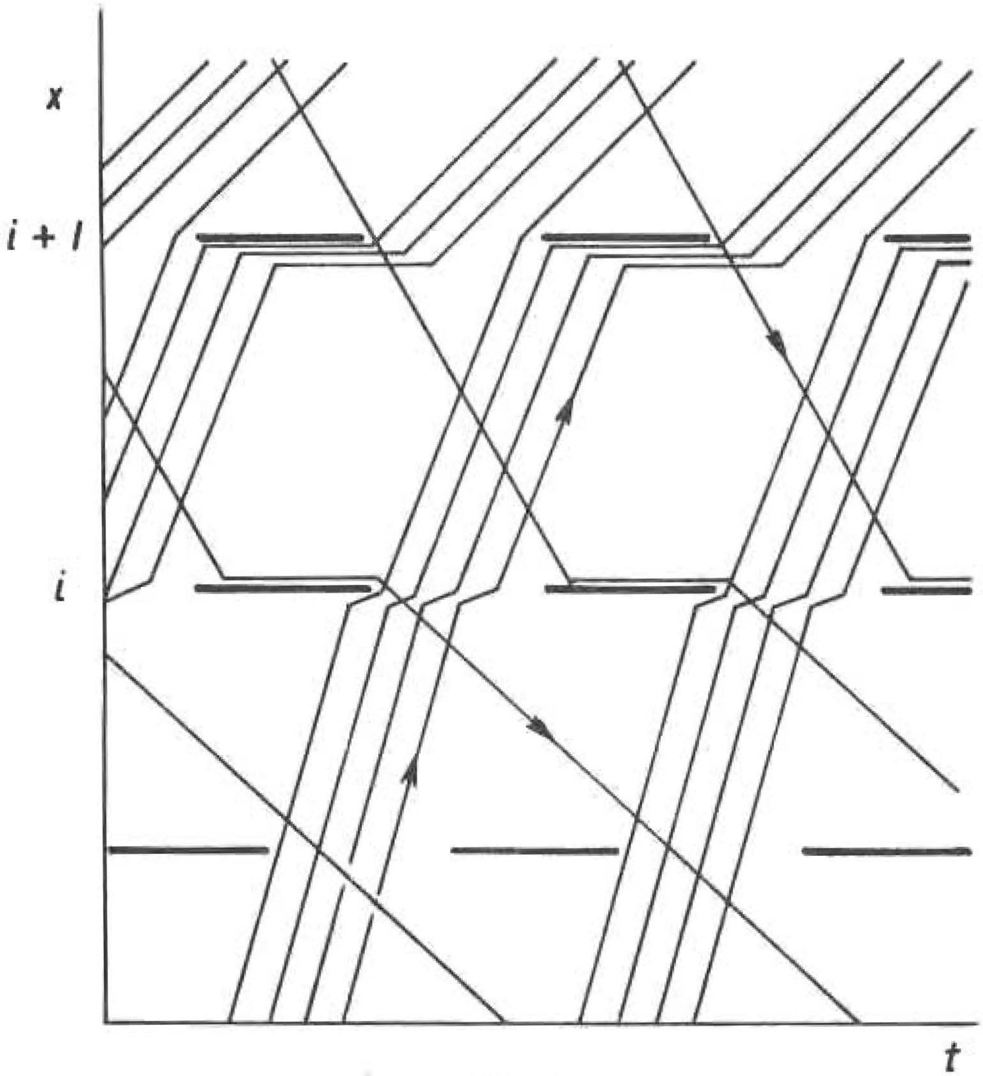

Fig. 1

If τ_i is the travel time from any intersection, i, to the next i+1 intersection, and τ_i' is the travel time from intersection i+1 to i, $\tau_i \neq \tau_i'$, we can make some elementary transformations which in effect make $\tau_i = \tau_i'$. Suppose we measure times at intersections i and i+1 with different clocks which are set so that the time t_i as seen by an observer at intersection i differs from that seen by an observer at intersection i+1 through

$$t_{i+1} = t_i - (\tau_i - \tau_i')/2.$$

A car traveling from i to i+1 starting at $t_i = 0$ for example will arrive at i+1 at time $t_i = \tau_i$ or

$$t_{i+1} = \tau_i - (\tau_i - \tau_i')/2 = (\tau_i + \tau_i')/2.$$

A car traveling from i+1 to i starting at $t_{i+1} = 0$ will arrive at i at time $t_{i+1} = \tau_i'$ or

$$t_i = \tau_i' + (\tau_i - \tau_i')/2 = (\tau_i + \tau_i')/2.$$

Thus by measuring times at various signals relative to nonsynchronous time origins, the original problem maps into one with equal "travel times" for the two directions between each pair i, i+1. Hereafter, we will choose $\tau_i = \tau_i'$.

The distances, the velocities of cars, and the travel times τ_i between different intersections may all vary with i, but the delays depend only on the τ_i. In Fig. 1, it is immaterial how we measure the vertical spacial coordinate as long as the diagram shows the correct τ_i. Instead of plotting actual distances we could just as well space the intersections i and i+1 a "distance" $v\tau_i$ apart, with v any arbitrary constant. The "trajectories" of cars would then have a "velocity" $\pm$ v between every pair of intersections.

Without loss of generality, we may therefore represent graphically any signal synchronization problem for which there is no spreading of platoons by an equivalent problem in which all cars travel with velocity v in one direction and -v in the other direction.

NEARLY SATURATED FLOWS

One of the complications associated with the analysis of traffic flow through a signal system arises from the possibility that a platoon of cars which leaves intersection i within a time less than the green time may be split into two parts at intersection i+1, part clearing at the end of a green and the rest at the start of the next green. The output from i+1 may, therefore, contain two disconnected platoons, one passing near the beginning and the other near the end of the green. These platoons may in turn be split into smaller pieces at other downstream intersections. According to the present idealized model with no platoon spreading, the flow from some intersection could contain arbitrarily many disconnected time intervals of flow within each green time. The number, widths, etc. of these flow intervals at intersection i will depend upon the total traffic flow, and the on the off-sets between signal i and arbitrarily many signals j, $j < i$. These long-range effects of offsets upon the flow pattern are particularly pronounced at low flows. Indeed the bandwidth for a system of signals depends upon the relative offsets between all signals, not just neighboring ones. For high flows, however, the delays and stops at one intersection depend mainly upon the offsets of nearby intersections and the analysis of the flow pattern becomes relatively simple.

Previously Buckley et al[5] considered the delay to a platoon traveling between only a pair of intersections, taking account of the fact that the platoon might be split into two parts at the downstream intersection. That the synchronization problem reduces to a study only of nearest neighbor intersections at nearly saturated flows was noted previously by the author[6], but in this previous paper the traffic was treated by car-following type models (with spreading of the platoon). Because of the complexity of the model, the consequences of this were not analized in detail.

In the following analysis we assume that all signals operate on a common cycle time T ; they all have the same green time G and the same capacity (sG in one direction and s'G in the other direction); and the flows through the system are arbitrarily close to these capacities (flows sG and s'G per time T in the two directions).

The flow leaving any signal is represented by a flow rate 0 during the red time R and a saturation flow rate during green time G. The most significant feature of this is that this flow pattern depends upon the time at which a cycle starts relative to some arbitrary time origin but is independent of the offsets of any other signals. The delay per cycle suffered at intersection i+1 by cars traveling from intersection i to i+1 depends upon the offsets between signals i and i+1, but is independent of offsets between any other signals. Also the delay per cycle suffered at intersection i by cars traveling from intersection i+1 to i depends upon the offsets only between signals i and i+1. The same properties are valid for the number of stops per cycle.

The total delay per cycle (or the total number of stops per cycle) in a system of intersections is the sum of the delays on all approaches to all intersections. If we consider only the delays to the through-traffic on the one street (not side-street delays), the delays (or stops) per cycle in the system can be written in the form

$$\text{delays (stops)} = \sum_i (d_i + d_i^*)$$

d_i = delay (stops) at signal i+1 for traffic moving from i to i+1

d_i^* = delay (stops) at signal i for traffic moving from i+1 to i.

The contribution for each i in the sum depends upon the offset between i and i+1 but not on any other offsets. The offsets which minimize the total delay (stops) are therefore those which minimize separately the contribution for each i. The global optimization problem thus reduces to a collection of separate problems for each pair of adjacent intersections (note that this is not the case for the bandwidth maximization).

To minimize delays or stops, it suffices now to study in detail only the delays between adjacent intersections.

ADJACENT INTERSECTIONS

For some pair of adjacent intersections i and i+1, let τ be the travel time from i to i+1 (or i+1 to i) and let δ be the offset of the signal i+1 relative to i. Since we will be considering here only a single pair of intersections, we will discard the implied index i on τ or δ. Also since any functions of τ and δ to be evaluated here will be periodic in time with period T, it will be convenient to agree that the values of τ, δ, or $\tau-\delta$ shall be interpreted as their values modulo T (or suitable equivalents).

Figs. 2a and 2b show two typical types of flow patterns from i to i+1. The broken lines represent the trajectories of cars leaving intersection i at either the beginning or the end of a green. If $0 < \delta - \tau < R$ as in Fig. 2a, then every car is delayed at i+1 by the same amount, $\delta-\tau$, and

$$\left.\begin{array}{ll} \text{total delay per cycle} & = sG(\delta-\tau) \\ \text{average delay per car} & = \delta-\tau \\ \text{maximum delay to any car} & = \delta-\tau \\ \text{average number of stops per car} & = 1 \\ \text{total number of stops per cycle} & = sG \end{array}\right\} \quad \text{if } 0 < \delta-\tau < R \tag{1}$$

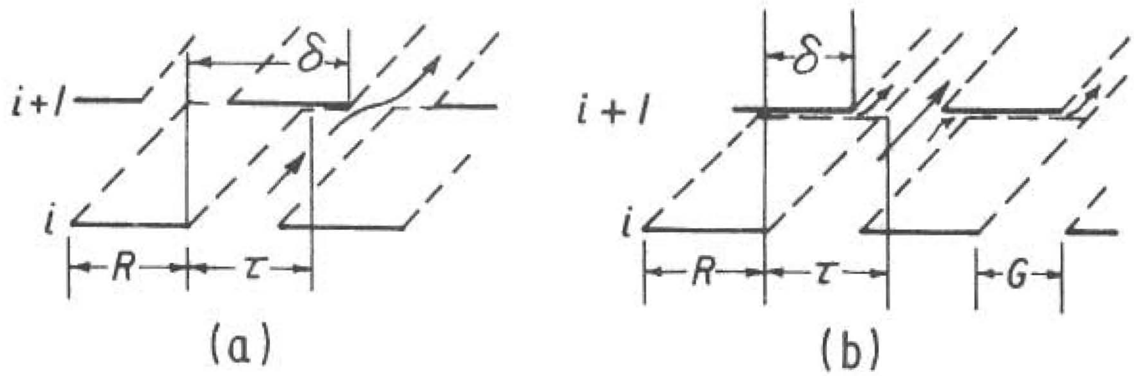

Fig. 2

If, however, $0 < \tau - \delta < G$ as in Fig. 2b, then those cars which leave from i+1 within a time $\tau - \delta$ after the signal at i+1 turns green have been delayed a time R. The remaining cars suffer no delay. Thus

$$\left.\begin{array}{lcl} \text{total delay per cycle} & = & (\tau - \delta)s\,R \\ \text{average delay per car} & = & (\tau - \delta)\,R/G \\ \text{maximum delay to any car} & = & R \\ \text{average number of stops per car} & = & (\tau - \delta)/G \\ \text{total number of stops} & = & (\tau - \delta)s \end{array}\right\} \quad \text{if } 0 < \tau - \delta < G \qquad (2)$$

Each of the above quantities considered as a function of δ is piecewise linear. The best offset is obviously $\delta = \tau$ for which no car is delayed; the worst is $\delta - \tau = R$ (or equivalently $\tau - \delta = G$) for which every car is delayed a time R. Fig. 3 shows some typical graphs of delay, stops, and maximum delay vs δ. These are shown only for $0 < \delta < T$ but are defined for other ranges of δ through the periodicity. Figs. 3b and 3c show that the objectives of few stops and of a small value for the largest delay to any car are not compatible with each other (except at $\tau = \delta$). If a signal stops only a few cars, these cars have a large delay (namely R) but if the signal causes a small maximum delay then all cars are stopped.

Equations (1) and (2) apply also for the traffic traveling from i+1 to i provided we reinterpret the parameters as those seen by traffic moving in the opposite direction. Both directions see the same values of G, R, and T. Also the travel time τ has been defined so as to be the same in both directions. The only changes in notation are that s is replaced by s' and the offset δ from i to i+1 is changed to the offset $-\delta$ from i+1 to i (or to $T - \delta$). The graphs of total delay or total stops vs. δ for the reverse traffic differ from those of Figs. 3a and 3b only in that the vertical scale is changed by a factor s'/s, and the graph is reflected about a vertical line at $\delta = T/2$ (equivalently δ is measured backwards from T instead of forward from 0). The analogue of Fig. 3c is obtained by only a reflection, but not a rescaling.

We are primarily concerned with the sum of the delays or stops for the two directions i. e. the

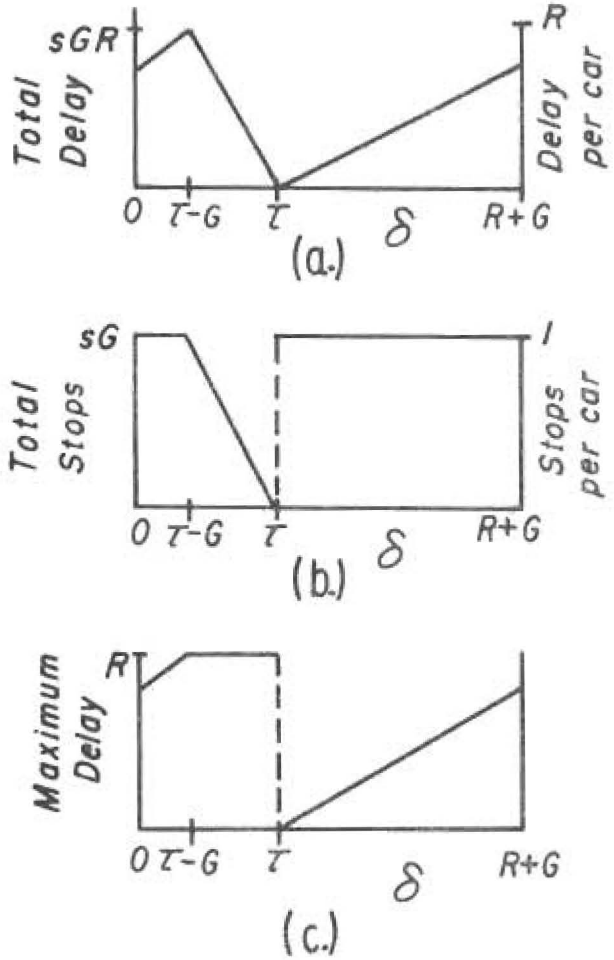

Fig. 3

sum of the graphs shown in Figs. 3a or 3b with their counterparts for the reverse flow. These composite curves are also piecewise linear but can have a variety of different geometries depending upon values of τ, R, G, and s'/s. These are illustrated in Figs. 4 and 5.

Fig. 4a shows some illustrations of total delay per cycle vs. δ. The broken line curves represent the delay at i+1 for the forward traffic, the dotted line the delay at i for the reverse traffic, and the solid line the sum of these two. In all cases the solid-line curves are piecewise linear with changes in slope at $\delta = \tau$, $\tau - G$, $-\tau$, and $G - \tau$. The location and even the order of these points depend upon the values of τ, G, and R, but not on s and s'. The slopes and heights of the curves, however, depend also on s and s'. Fig. 3a applies for $G < R$ and $s = s'$. We will examine this case first.

The choice of $s = s'$ simplifies the delay curves in that the delay as a function of δ (extended by periodicity outside $0 < \delta < T$) must be symmetric about $\delta = 0$ and $\delta = T/2$. Since the total delay curves are piecewise linear, this symmetry implies that the curves must have slope 0 at $\delta = 0$ and $\delta = T/2$, unless $\delta = 0$ or $T/2$ happens to be a point of discontinuity for the slope. It follows also from these properties that the minimum delay must be realized at either $\delta = 0$ or $\delta = T/2$, although a zero slope at these points implies that the values of δ which give minimum delay are not unique.

The top graph, (1), Fig. 4a for $\tau = 0$ (or any multiple of T) shows a unique setting $\delta = 0$ for minimum delay. In this case $\delta = 0$ is simultaneously the optimal for both directions. For $0 < \tau < G/2$ as in graph (2), the optimal setting is still at $\delta = 0$, but also for any δ with $-\tau < \delta < \tau$. The minimum delay is now positive. At $\tau = G/2$ in graph (3), the total delay is independent of δ. For $G/2 < \tau < G$ in graph (4), the optimal setting shifts to the range $\tau < \delta < T - \tau$ including the point $\delta = T/2$. As τ increases further the minimum delay at $\delta = T/2$ decreases. There are some variations in the shape of the delay curves at $\delta = 0$ as τ increases past G, but finally at $\tau = T/2$ the minimum delay again reaches zero when the half cycle offset is simultaneously optimal for both directions. As τ increases past T/2 the pattern of delay curves repeats itself but with a half cycle phase shift. Thus diagrams (7) and (8) are the same as (1) and (2) except for a shift in δ by T/2.

Although the curves of Fig. 4a apply only for $s = s'$ and $G < R$, it is easy to see from these diagrams what happens in more general cases. If, for example, $s > s'$ we simply add to the present type total delay curves an extra amount proportional to $(s - s')$ times the delay curve for the forward traffic (broken line). The horizontal portions of the composite delay curves will now be sloped in such a way that the optimal setting is always at $\delta = \tau$, the setting which gives zero delay to the heavier traffic.

The case $G > R$ is only slightly different from $G < R$. If in Fig. 3a the role of G and R were interchanged sending $G < R$ into $G > R$, the graphs for delay would be reflected about the line $\delta = \tau$ i. e., $\delta - \tau \rightarrow \tau - \delta$. In Fig. 4a, the broken line curves would map into curves like the dotted line curves and vice-versa. The total delay would retain the same form as shown in Fig. 3a except that τ would be mapped into $-\tau$ (or $T - \tau$). Fig. 5a shows the resulting sequence of diagrams analogous to Fig. 4a but for $G > R$.

In any case, the delay will be at least for $\delta = \tau$ if $s > s'$ and for $\delta = -\tau$ if $s < s'$. If s is close to s', the delay vs. δ curves usually will be quite flat near the minimum. If $s = s'$ the optimal value of δ is typically not unique. From the various cases shown in Figs. 4a and 5a, plus their extensions to $T/2 < \tau < T$ one can construct Fig. 6a. For any value of τ, the values of δ contained in the shaded areas of the $\delta - \tau$ space of Fig. 6a represent those values which give the minimum delay if $s = s'$. For example, if $0 < \tau < G/2$, the optimal values of δ are $0 \leq \delta \leq \tau$ and $T - \delta \leq \delta \leq T$. The same figure applies for either $G < R$ or $R < G$.

Figs. 4b and 5b for $s = s'$ show the total number of stops per cycle for the same sequence of τ values as in Figs. 4a and 4b. The shapes of these curves are slightly different for $G < R$ and $G > R$ but have the following common features. For $\tau = 0$ the number of stops is zero at $\delta = 0$ but is discontinuous at $\delta = 0$. For $0 < \tau < G/2$ the total stops has a minimum for $-\tau < \delta < \tau$, the same as for the delay. For $G/2 < \tau < T/2$, however, the minimum stops occur at $\delta = \tau$ and $\delta = T - \tau$ only, the optimal setting for one direction or the other. At $\tau = T/2$, the total stops drops to zero again discontinuously at $\delta = T/2$. The pattern then repeats for $\tau > T/2$ except with a shift of δ by $T/2$.

If $s > s'$, the curves for total number of stops differ from those of Figs. 4b and 5b in that an excess of the broken line curves must be added. This will in all cases cause the minimum number of stops to occur at $\delta = \tau$ (uniquely) favoring the direction of traffic with the heavier flow.

Fig. 6b shows the values of τ which give the minimum number of stops for $s = s'$. For $0 < \tau < G/2$ or $T/2 < \tau < T/2 + G/2$, the settings which give minimum delay are also those which give minimum stops. For $G/2 < \tau < T/2$ or $T/2 + G/2 < \tau < T$, however, the settings which give minimum stops are only $\delta = \pm \tau$ whereas those which give minimum delay cover a range of values between these.

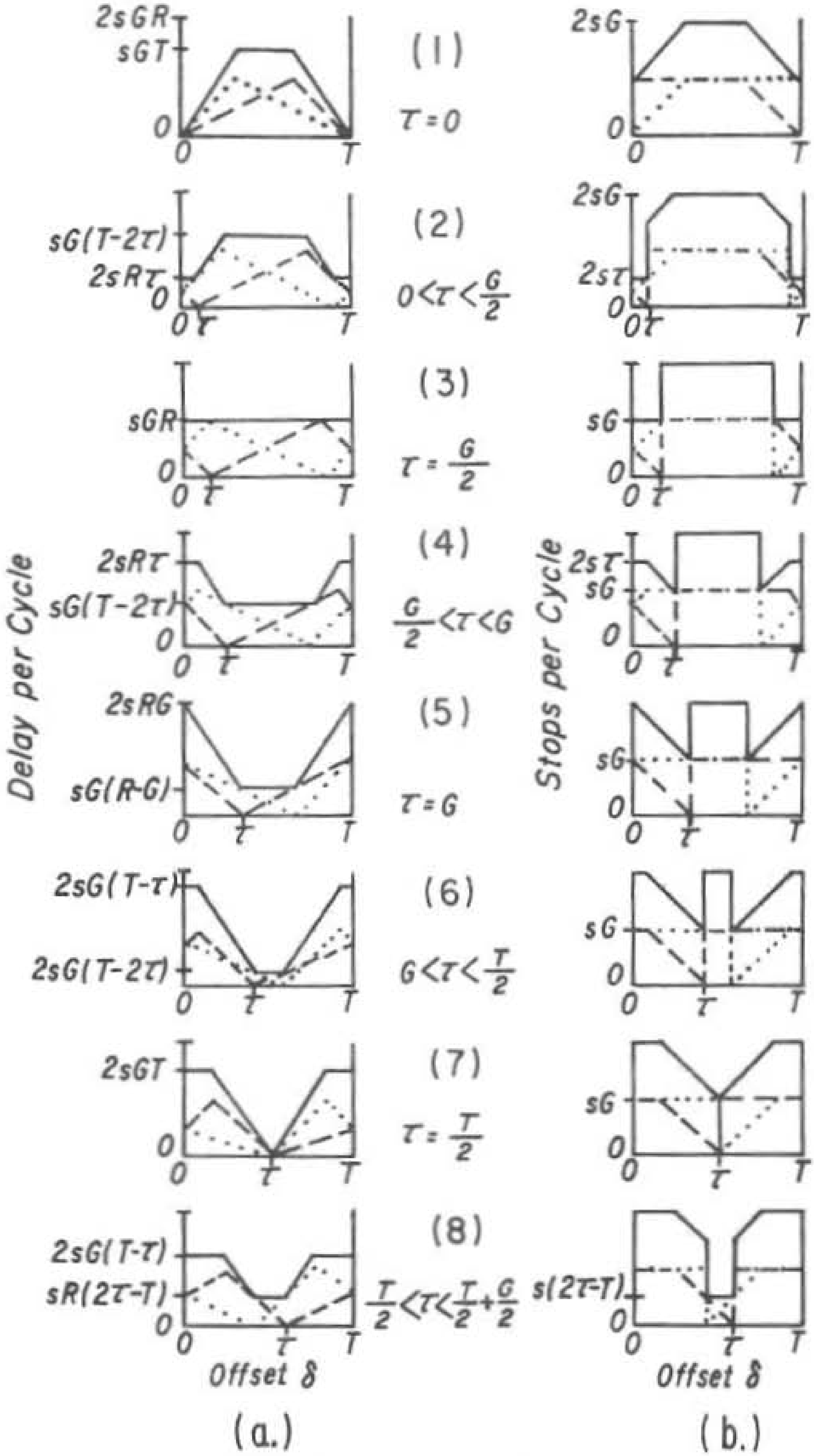

Fig. 4 — G less than R.

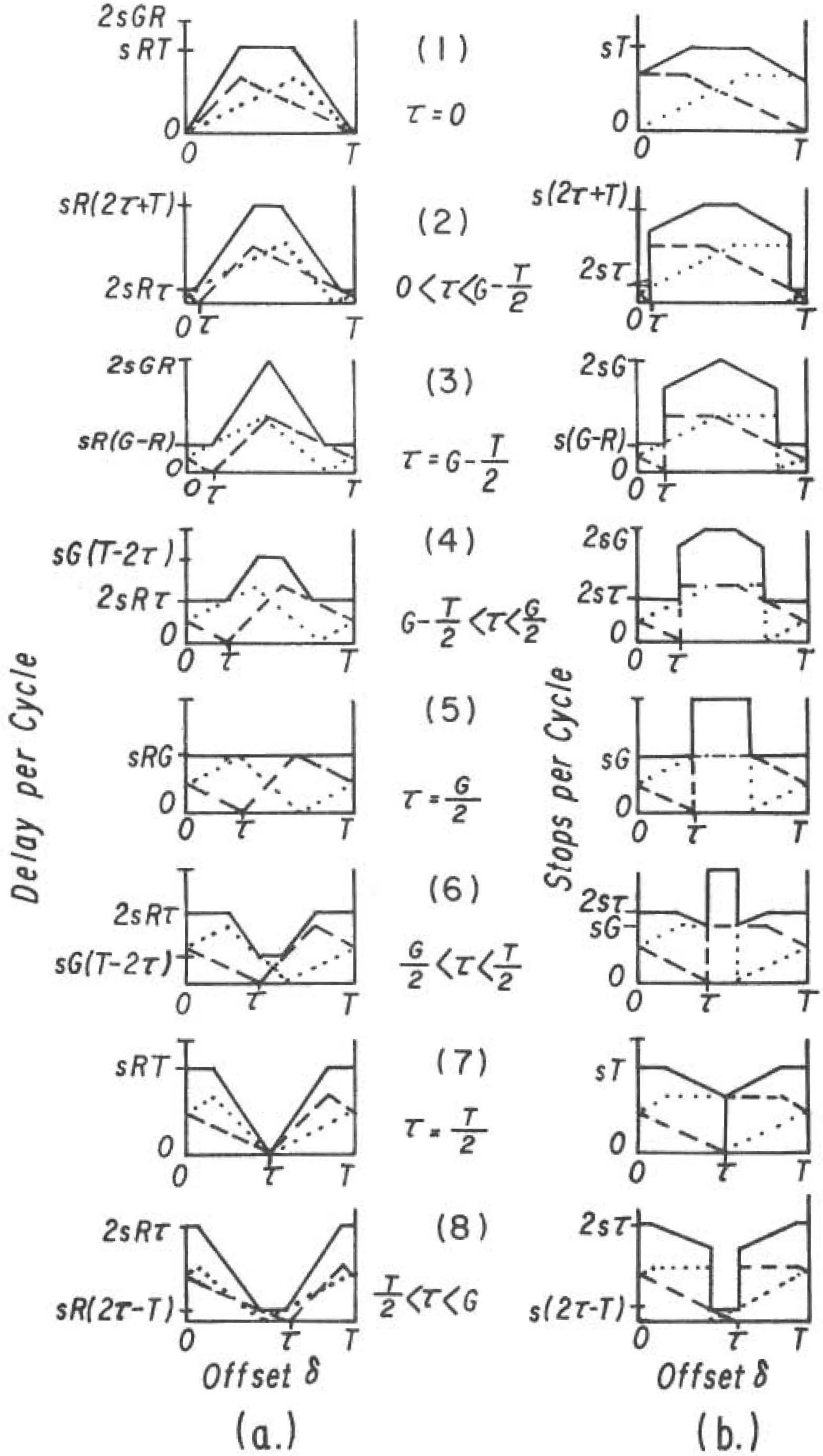

Fig. 5 — G greater than R.

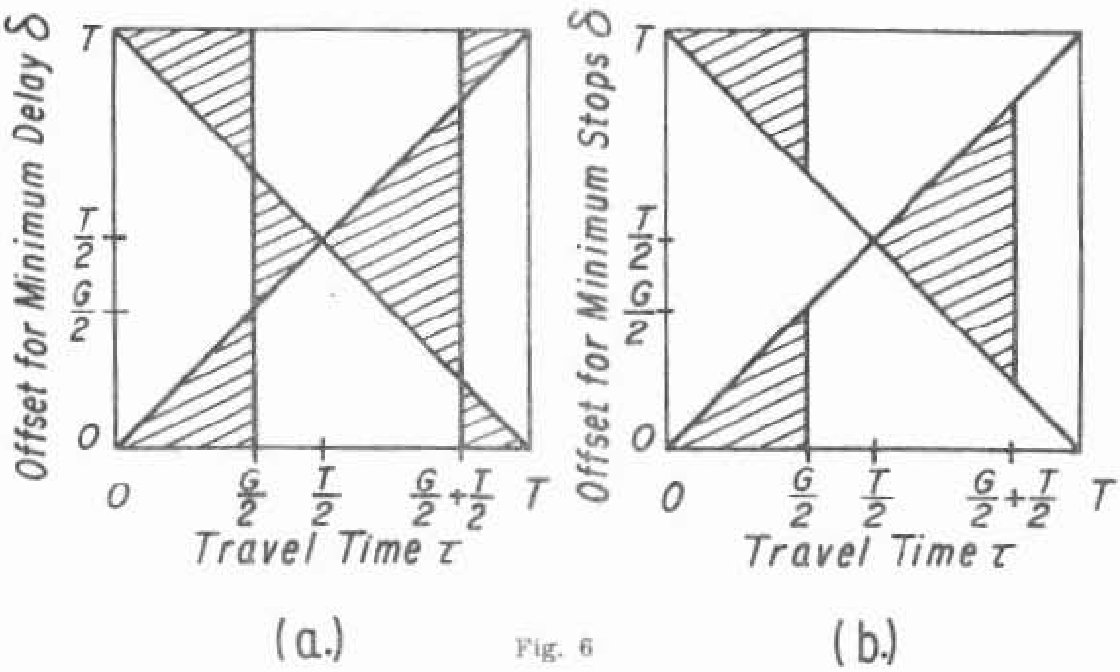

Fig. 6

INTERPRETATION

The model described here is severely restricted by the assumptions that (1) there is no platoon spreading, (2) there is no turning traffic, and (3) the flow is nearly saturated. Within these limitation, however, we conclude that it is possible to simultaneously minimize the total delay and the total number of stops at all intersections for the through traffic. A sufficient condition for this is that the offsets give a one-way progression for the direction of traffic with the larger flow ($\delta_i = \tau_i$ for all i, or $\delta_i = -\tau_i$ for all i). If the flows are equal, however, there are other optimal settings, and if the flows are nearly equal there may be a wide range of settings which give nearly minimal delays and stops. In the latter case, the most desirable setting within this range of nearly optimal settings may be determined by conditions other than delays or stops, or may be modified considerably by more accurate estimates.

Some of the objections to a one-way progression on a two-way street are: (1) the delays are unequally distributed between the two directions of flow, (2) if the street in question is part of a network, the progression may conflict with the optimal settings for other streets, (3) if there is turning traffic, these settings cause large delays to traffic turning into the main street progression from the side street, and (4) since travel times vary during the day, the settings must also change during the day.

Suppose, for example, that traffic is so heavy that one must employ a long cycle time to achieve adequate capacity, causing the green time to be larger than $2\tau_i$ for all i. If, in addition, the flows are nearly equal in the two directions, the above theory would seem to justify the commonly used scheme of zero offsets ($\delta_i = 0$ for all i). From Figs. 6a and 6b we see that for $\tau_i < G/2$ any offset with $-\tau_i \leq \delta \leq +\tau_i$ gives both the minimum stops and delay. Which of these is most desirable, therefore, must be based upon other considerations. The choice $\delta_i = 0$ for all i has the advantage that it (1) gives equal delays to the traffic in both directions, (2) is compatible with a network synchronization with $\delta = 0$ also on all other streets, (3) is not obviously bad for the turning traffic, and (4) does not depend upon the speed of traffic.

If the traffic is heavy in both directions, it is certainly desirable to set the signals so that all drivers traversing the entire sequence of signals suffer comparable delays regardless of direction. Since by our hypothesis, cars traveling in the same direction all have the same velocity, the order in which cars leave one intersection is the same as for any other intersection. All cars which leave the first intersection within the same green and also leave the last intersection within the same green have identical travel times. Those which leave the last intersections during different green times have a difference of travel time equal to R, independent of how many signals there may be in the system. For a long sequence of signals these differences of R in delays for cars traveling in the same direction are not of great practical importance. Cars traveling in opposite directions could, however, have widely different delays depending upon the number of signals and the scheme of synchronization.

In contrast with the arguments of the last section which were concerned with the total delay to all cars at a single intersection (and the total delay to all cars at all intersections), the present argument deals with the total delay to a single car at all intersections. Even though different cars may suffer different delays at different intersections, the above conclusions imply that it is not possible for the differences in delays between two cars traveling in the same direction to accumulate. A driver who suffers a relatively large delay at one intersection will find himself in a favored position at some later intersection.

If $s = s'$, then regardless of the values of the τ_i, it is always possible to set a sequence of signals so as to simultaneously minimize total stops and total delay, and at the same time keep the difference in delays between any two cars less than R, even if the two cars travel in opposite directions. If, for example, the cumulative delay at all signals to a car traveling in the forward direction ($i \rightarrow i+1 \rightarrow i+2, \ldots$) is less than that in the reverse direction, then any signal having a δ_i which favors the traffic in the forward direction ($\delta_i = \tau_i$ for example) can be switched to $-\delta_i$ without causing any change in the total delays or stops for the entire system. This will, however, increase the travel times in the forward direction and decrease those in the reverse direction. This can be done for as many i values as necessary to bring the asymmetry in travel time to a point where any additional changes will only switch the favor to the opposite direction. One possible scheme would be to set several consecutive δ_i at $\delta_i = \tau_i$, then set the next several δ_i at $\delta_i = -\tau_i$ etc. balancing the number of signals with $\delta_i = \tau_i$ and $-\tau_i$ so as to keep the delays in the two directions nearly equal. There is obviously still a considerable number of ways in which these objectives can be met.

If $q > q'$, however, the settings which minimize delays and stops are unique and favor the forward traffic ($\delta_i = \tau_i$ for all i). There is no setting which simultaneously minimizes both stops and delays and gives also similar travel times for the two directions (unless the τ_i are all 0 or $T/2$). If we assign some economic values to delays, stops and asymmetry in travel time, we can, however, find settings which minimize cost by compromising these conflicting objectives. We may for example set $\delta_i = \tau_i$ for most but not all i.

Regarding the network synchronization problem, it is clear that a scheme of optimal synchronization for one set of streets in a network may be incompatible with the optimal for other streets, particularly if the streets carry one-way traffic. The theory described here suggests, however, that the conflict in objectives between traffic moving in opposite directions on the same street is typically much more severe. Since the delays and stops for nearly equal flows on a two-way street are rather insensitive to the settings over a wide range of values, it is not unlikely that one should find settings which minimize delays and stops simultaneously on all streets (but not for both directions on the same street). Thus, for example, the setting $\delta_i = 0$ is optimal in many situations.

We have not discussed here the problem of choice in cycle time. Obviously the delays are sensitive to the cycle time and traffic engineers do typically select a cycle time so as to make τ_i a multiple of T/2 for some i (if possible). If the τ_i are different between different intersections, it is, of course, not usually possible to do this for all i. Sometimes it is not possible in practice to do this for any i. If the intersections are close together, the time $T = 2\tau_i$ may be too short to give the necessary capacity. The choice of T is certainly an important aspect of the problem, but a change in cycle time will also change the capacity. Since the present theory is restricted to flows per cycle equal to the capacity per cycle, an increase in cycle time is likely to change the problem into one outside the scope of the present analysis.

REFERENCES

1. Baverez, E. and G. F. Newell, "Traffic Signal Synchronization on a One-way Street," Transportation Science 1, 1967.

2. Morgan, J. T. and J. D. C. Little, "Synchronizing Traffic Signals for Maximal Bandwidth," Operations Research 12, 1964, p. 896-912.

3. Little, J. D. C., B. V. Martin, and J. T. Morgan, "Synchronizing Traffic Signals for Maximum Bandwidth," Highway Research Record #118, 1966, p. 21-47.

4. Little, J. D. C., "The Synchronization of Traffic Signals by Mixed-Integer Linear Programming," Operations Research, 14, 1966, p. 568-594.

5. Buckley, D. J., J. A. Tomlin, and W. G. a'B. Minson, "Delay to Traffic Platoons," Proceedings, Third Conference of the Australian Road Research Board, 1966.

6. Newell, G. F., "Synchronization of Traffic Lights for High Flow," Quarterly Applied Mathematics 21, 1964, p. 315-325.