Lost Time and Cycle Length for Actuated Traffic Signal

Peter G. Furth, Burak Cesme, and Theo H. J. Muller

Time within an actuated signal cycle can be decomposed into time that is fully used, which is the saturation headway multiplied by the number of passing vehicles, and time that is wasted or lost. Activity network modeling is used to show the interaction between signal timing events and traffic flow transitions. Seven components of generalized lost time are identified: those associated with start-up, minimum green, parallel queue discharge (for simultaneous gap-out), extension green, parallel extension (for nonsimultaneous gap-out), the passing of the critical gap, and phase end. Simple formulas can be used to estimate all of these components for many practical cases, allowing one to estimate average cycle length without iteration. The modeling framework accounts for the dual-ring structure with minimum green and maximum green constraints and on–off settings for recall and simultaneous gap-out. Experiments with microsimulation software verify the formulas developed. The formulas show the sensitivity of lost time, and therefore average cycle length, to parameters that a designer can control including detector setback, critical gap, gap-out settings, and number of lanes. They also show sensitivity to total demand and to the ratio of noncritical to critical phase volumes.

At a fully actuated traffic signal, cycle length varies from cycle to cycle as an outcome of traffic demand and various physical and timing parameters. Average cycle length is an important measure of performance of an actuated signal because average delay for motorists and pedestrians is roughly proportional to cycle length. Traffic engineers need more transparent tools for predicting average cycle length in order to better understand the relationship of signal design to performance. Microsimulation software offers one such tool; however, many traffic engineers lack a facility with microsimulation and would benefit from an approach that is intuitive and simple.

An activity network and a generalized concept of lost time are used to model the operation of a fully actuated signal with a dual-ring structure. The focus here is on time in a signal cycle that is wasted (lost) as a basis for determining cycle length, and seven components of lost time are identified. Compared with published methods of cycle length estimation, the model discussed here represents some features of actuated operation more accurately, such as joint gap-out logic and minimum green. It is shown that in some common situations, the expected cycle length cab be determined without iterative calculation.

P. G. Furth and B. Cesme, Department of Civil and Environmental Engineering, Northeastern University, 360 Huntington Avenue, Room 400 SN, Boston, MA 02115. T. H. J. Muller, Transportation and Planning Department, Delft University of Technology, Stevinweg 1, Delft 2628 CN, Netherlands. Corresponding author: P. G. Furth, pfurth@coe.neu.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2128, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 152–160.
DOI: 10.3141/2128-16

EXISTING CYCLE LENGTH PREDICTION METHODS

The seminal research on actuated operation by Newell and Osuma (*1*) showed that although a pair of one-way traffic streams could be easily analyzed and efficiently operated, complications arising when parallel streams were considered made the analysis muddy and removed much of the theoretical efficiency of actuated control. Their analysis assumed that controllers can detect the moment of queue discharge when in fact the performance of actuated controllers is driven by their inability to discern a sharp boundary between saturated and unsaturated flow.

Lin (*2*) considered multiphase control with the gap-seeking logic of modern controllers together with their minimum and maximum green constraints. His work focuses on estimating the length of the extension green period and, for that purpose, modeling headway distributions for single and multilane approaches. He proposes an iterative, deterministic approach to estimating cycle length without considering interactions with noncritical phases. Akcelik (*3*) built on Lin's work as part of NCHRP Project 3-48; since 1997, his work has been adopted in the *Highway Capacity Manual* (HCM) (*4*). It includes further development of headway distributions and consideration of dual-ring control. However, interaction with noncritical phases is almost trivial because like Lin, a largely deterministic analysis is used, which predicts that noncritical phases never affect cycle length.

LOST TIME IN CRITICAL CIRCUIT

A four-leg intersection is considered with left and through–right phases on each approach, with signal control following the standard dual-ring, eight-phase structure, as shown in Figure 1. Phase lengths (splits) vary from cycle to cycle depending on traffic. Barriers in the dual-ring structure divide the signal cycle into half-cycles, each containing two half-rings that must start and end simultaneously at the barriers but otherwise can run independently. In any given pass through a half-cycle, depending on traffic, either half-ring may be dominant. The half-ring with the greater average demand is considered the critical half-ring, and its phases critical phases. Because of the barriers, the cycle length can be treated as the sum of the times (splits) of the four critical phases.

For each critical phase, the time is divided into used and lost time. For every vehicle that passes the stopline, a time ($1/s^i$), the saturation headway for phase *i*, is accounted as used; the remainder of the split is treated as lost time. Thus, if the saturation headway is 2 s and two cars pass with a headway of 3 s, 2 s is accounted as used and 1 s as lost. With the familiar terms L^i, total lost time for phase *i*; v^i, approach volume for phase *i*; s^i, saturation flow rate for phase *i*; and *C*, cycle length and use of expected values of *C* and L^i, the mean number of phase *i* vehicles served per cycle is Cv^i, and so, summing over critical

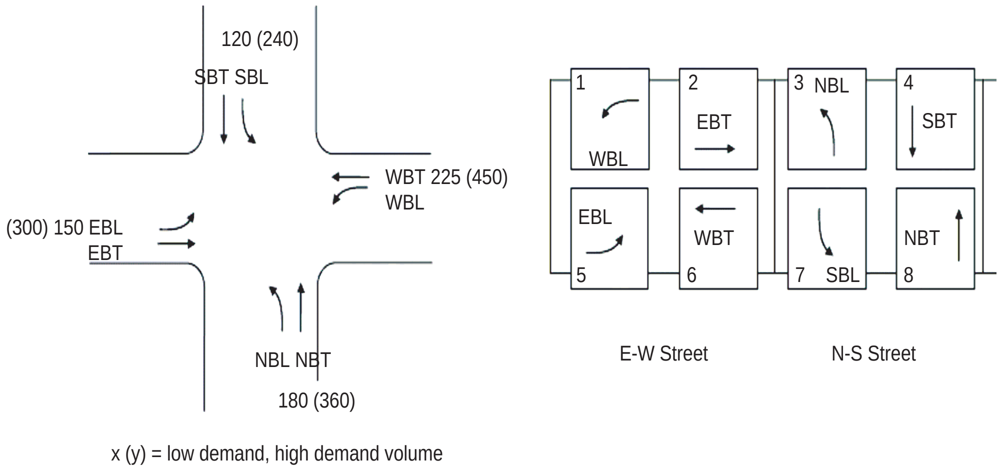

FIGURE 1 Example intersection with eight-phase dual-ring control (SBT = southbound through, SBL = southbound left, WBT = westbound through, WBL = westbound left, NBL = northbound left, NBT = northbound through, EBT = eastbound through, and EBL = eastbound left).

phases, the mean amount of time used per cycle is $\Sigma Cv^i/s^i$. At the same time, the mean amount of time used per cycle is $C - \Sigma L^i$. Equating these quantities leads to the familiar cycle time formula:

$$C = \frac{\sum_{i\text{critical}} L^i}{1 - \sum_{i\text{critical}} \left(\frac{v}{s}\right)^i} \tag{1}$$

The usefulness of Equation 1 depends on an ability to determine generalized lost time.

ACTIVITY NETWORK FOR PHASE NOT ENDING AT BARRIER

An actuated signal cycle can be modeled as an activity network (*5, 6*), in which each node represents a moment in time and each arc represents an interval. Figure 2 shows an activity network for a single phase that terminates independently of other phases (i.e., does not end at a barrier). The convention that square nodes represent events in the usual sense of the word is used, with the following meanings:

SG = start of green (start of split);
SQ = start of queue discharge;
EQ = end of queue discharge;
SX = start of extension green;
UX = start of unsaturated part of green extension;
LXD = moment at which front of last vehicle that extends green passes downstream edge of extension detector;
LXS = moment at which front of last vehicle that extends green passes stopline;
GD = moment at which gap long enough to end green is detected (gap-out);
SY = start of yellow;
SPX = start of parallel extension flow period (which appears only in Figure 3);
SLA = start of late arrival period; when cars that arrive after gap-out but before clearance time are served;
SCl = start of effective clearance time;
E = end of split; and
MaxG = maximum green expired (max-out).

Round nodes are not events in the usual sense of the word; they are a device used to divide an interval into subintervals of interest.

The network has three kinds of arcs:

- Solid lines represent intervals with externally determined length;
- Dashed lines represent slack time that ensures that every arc arriving at a common node arrives at the same time; the length of slack arcs must be nonnegative; when two slack arcs arrive at a common node, one of them usually has zero length in any given pass through

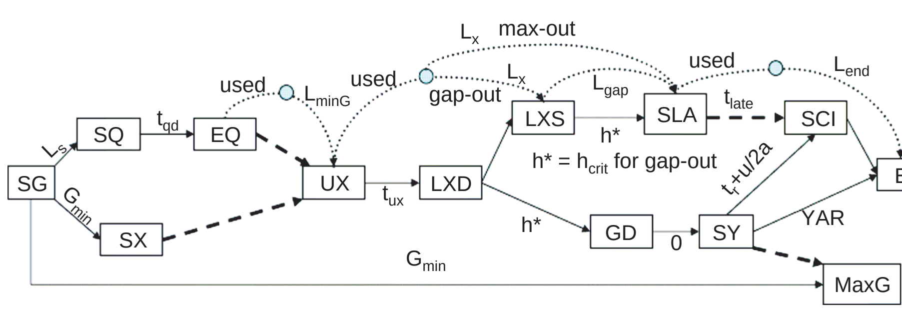

FIGURE 2 Activity network for phase not ending at barrier.

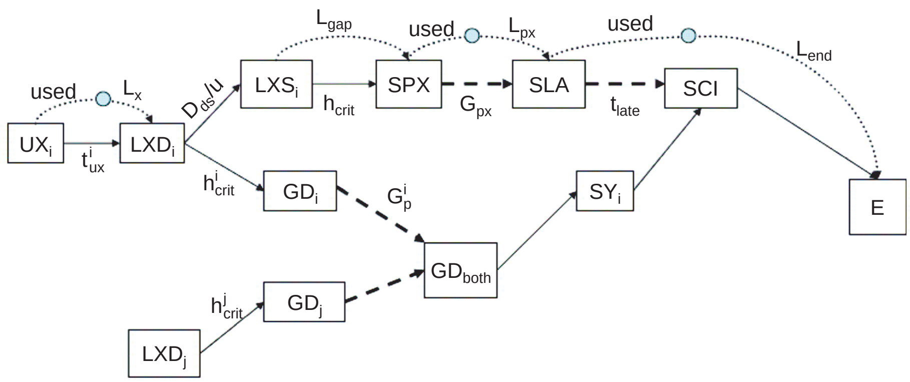

FIGURE 3 Nonsimultaneous gap-out.

the cycle; however, averaging over many cycles, both may have a nonzero expected length; and

- Dotted lines represent subdivisions of an interval into used time and lost time.

Activities and events concerned with signal timing are shown in the lower part of Figure 2, and activities and events concerned with traffic flow are shown in the upper part. In both the network and the equations that follow, phase indices are suppressed except where phases interact.

Start-Up Lost Time, Queue Discharge, and Minimum Green

In the upper part of Figure 2, arc SG–SQ represents start-up lost time with length L_s, as defined in the HCM. Arc SQ–EQ represents queue discharge (t_{qd}) at a uniform rate *s.* Traffic volumes and maximum green settings are assumed to be such that an approach can discharge its queue before reaching maximum green. Expected queue discharge time is given by the following well-known formula:

$$t_{\text{SQ-EQ}} = \frac{rv}{(s-v)}$$

where *r* is the effective red, *v* is the arrival volume, and *s* is the saturation flow rate.

In the lower part of Figure 2, arc SG–SX represents a phase's minimum initial green period $G_{\min}$, after which the phase enters its extension green period. Controllers have a minimum green period for one or all of the following purposes: to allow traffic flow to get beyond start-up irregularities, to prevent gap-out until after the discharge of a queue that does not reach back as far as an upstream detector, and to provide sufficient pedestrian crossing time.

UX, the start of the unsaturated part of extension green, has two predecessor events connected to it by slack arcs: with respect to timing, UX follows Event SX, and with respect to traffic flow, it follows Event EQ. When EQ precedes SX, a condition called early queue discharge, the minimum green constraint governs. In that case, the slack arc from EQ to UX has nonzero length. If *n* is the number of arrivals during the preceding red period, early queue discharge will occur when

$$n < n_0$$

with

$$n_0 = (s-v)(G_{\min} - L_s)\,(\text{early queue discharge}) \tag{2}$$

in which case the unsaturated part of the minimum green has the following length:

$$t_{\text{EQ-UX}} \,|\, n = \frac{G_{\min} - L_s - n}{(s-v)} \qquad \text{for } n < n_0 \tag{3}$$

During this slack period, cars are expected to arrive at a rate *v,* and each arrival will use a period 1/*s.* The unused part of this slack period is $L_{\min G}$, lost time due to minimum green:

$$L_{\min G} \,|\, n = (G_{\min} - L_s)\left(1 - \frac{v}{s}\right) - \frac{n}{s} \qquad n < n_0 \tag{4}$$

Taking expectation over possible values of *n,*

$$L_{\min G} = P(n < n_0)\left[(G_{\min} - L_s)\left(1 - \frac{v}{s}\right) - \frac{E(n \mid n < n_0)}{s}\right] \tag{5}$$

Equation 5 applies when the phase is set to recall, meaning that it may not be skipped. If the phase is not set to recall, it will be skipped if $n = 0$, in which case there will be no lost time. For a phase without recall,

$$L_{\min G} = P(0 < n < n_0)\left[(G_{\min} - L_s)\left(1 - \frac{v}{s}\right) - \frac{E(n \mid 0 < n < n_0)}{s}\right] \tag{6}$$

Akcelik (*3*) and the HCM (*4*) use a deterministic approach to model the impact of the minimum green. For a given cycle length, they determine $E[n]$, and depending on whether it is less than n_o, minimum green is assumed to govern either in every cycle or never. In spite of being iterative, this approach underestimates the impact of minimum green on cycle length for phases in which minimum green governs in some, but not all, cycles. If it is assumed that minimum green always governs, there is an error from ignoring those cycles in which the phase lasts longer than the minimum green, and if the minimum green constraint is ignored, there is an error for those cycles in which it actually extends the phase length.

Extension Green and Extension Lost Time

During the extension green, if the maximum green constraint is not binding, the signal is kept green until the Event GD, when a gap longer than a controller parameter minGap is detected by a presence detector. This parameter is invariably set long enough that a phase will not gap out during queue discharge. minGap corresponds to a critical headway:

$$h_{\text{crit}} = \min \text{Gap} + \frac{(\text{Len}_d + \text{Len}_v)}{u} \tag{7}$$

where

Len_d = detector length,
Len_v = vehicle length, and
u = approach speed in the absence of queues.

Using h_{crit} rather than minGap allows one to treat detection as an instantaneous event occurring when the front of a vehicle passes the downstream edge of a detector.

The time during which traffic flows between SQ and SCl can be divided into four periods with differing flow rates:

- Queue discharge with flow rate *s,* ending at Node EQ (which is simultaneous with Node UX when minimum green is not binding);
- An unsaturated flow period during which headways are subcritical—that is, shorter than h_{crit}—ending with Events LXD and LXS when the last extending vehicle (last vehicle with subcritical headway) passes the detector and the stopline, respectively;
- Period of no flow while the critical headway passes, ending with Event SLA; and
- Late arrival period at flow rate *v* running from SLA to SCl.

For a phase that terminates independent of other phases, the same general logic is followed as that of Akcelik and the HCM for determining the duration of the unsaturated extension green. Headways are assumed to be random and independent. The unsaturated flow period consists of a series of subcritical headways. On the basis of the chosen headway distribution, one can determine $p = P(H < h_{\text{crit}})$ and $E(H | H < h_{\text{crit}})$, where H is headway length. Headways are treated as a sequence of Bernoulli trials, with the expected number of headways before gap-out equal to

$$n_{\text{ux}} = \frac{p}{(1-p)} \tag{8}$$

(If a phase is prone to sometimes max out, that is, end its green because of the maximum green constraint, Equation 8 overestimates n_{ux}.)

Multiplying by $E(H | H < h_{\text{crit}})$ gives t_{ux}, the length of the unsaturated extension period, and subtracting the used time (n_{ux}/s) gives the extension lost time, L_x:

$$L_x = n_{\text{ux}} * \left(E[H | H < h_{\text{crit}}] - \frac{1}{s} \right) \tag{9}$$

In the activity network in Figure 2, a distinction is made between when the last green-extending vehicle is detected (LXD) and when it crosses the stopline (LXS). In intersection flow models, arrivals are defined with reference to the stopline. Therefore, the time covered by the n_{ux} subcritical headways runs from Node UX to Node LXS.

With a shorter, "snappier" critical gap, extension lost time is diminished by two mechanisms: p is shortened, reducing n_{ux}, and the lost time per subcritical headway is reduced as well.

Unsaturated Headway Distribution

The form of the assumed unsaturated headway distribution matters mostly with single-lane approaches on which very small headways do not occur. Akcelik and the HCM use a two-parameter modification of the exponential distribution. One parameter is $h_{\min}$, a minimum headway; the other is a bunching parameter that determines the proportion of headways that equal the minimum headway.

A special case of the two-parameter model has only one externally specified parameter, $h_{\min}$. It assumes an underlying exponential distribution with parameter λ and transforms to $h_{\min}$ all headways shorter than $h_{\min}$. With this distribution,

$$p_{\min} = P(H = h_{\min}) = 1 - \exp(-\lambda h_{\min}) \tag{10}$$

and

$$E[H] = h_{\min} p_{\min} + (1 - p_{\min})\left(h_{\min} + \frac{1}{\lambda}\right) \tag{11}$$

The value of λ is determined by equating $E[H]$ to $1/v$. Then

$$p = P(H < h_{\text{crit}}) = 1 - \exp(\lambda h_{\text{crit}}) \tag{12}$$

$$E[H | H \leq h_{\text{crit}}] = \frac{3{,}600}{pv} - \frac{1-p}{p}\left(h_{\text{crit}} + \frac{1}{\lambda}\right) \tag{13}$$

In numerical experiments with the simulation software VISSIM, the authors found that on one- and two-lane approaches, the one-parameter model yielded a slightly better fit than the two-parameter model with parameter values recommended in the HCM.

Gap Lost Time and End Lost Time

If the maximum green constraint is not binding, the final portion of a phase's split begins at LXD, when the last extending vehicle passes the detector. The signal timing track continues through Nodes GD and SY (which follows GD immediately for phases not ending at a barrier) and, after the yellow and all-red interval (YAR), Event E.

On the vehicle flow side, LXS follows LXD by the travel time D_{ds}/u, where D_{ds} is the distance from the downstream edge of the extension detector to the stopline and u is speed. If the downstream edge of the extension detector lies beyond the stopline, this travel time is negative. Following LXS there is a period of no flow of length h_{crit} as the critical gap passes, ending at Node SLA; this interval is a component of lost time called gap lost time, given by

$$L_{\text{gap}} | \text{gap-out} = h_{\text{crit}} \tag{14}$$

Equation 14 shows how the minimum gap setting, already shown to affect extension lost time in two ways, affects gap lost time as well.

The clearance interval begins at the moment within the YAR period after which vehicles no longer enter the intersection. Assuming that vehicles follow the "stop if you safely can" convention on seeing the yellow signal, Event SCl follows the onset of yellow by the interval

$$t_r + \frac{u}{2a} \quad (15)$$

where t_r is the reaction time, taken in these experiments to be 1.0 s, and a is the deceleration rate, taken to be 0.35 g = 11.3 ft/s^2.

From the network diagram, the length of the late arrival window is

$$t_{\text{late}} = \max\left(t_r + \frac{u}{2a} - \frac{D_{\text{ds}}}{u}, 0\right) \quad (16)$$

The expected number of late arrivals is vt_{late}, and the time they use is $(v/s)t_{\text{late}}$. End lost time represents the clearance interval and the unused part of the late arrival interval:

$$L_{\text{end}}\big|\text{gap-out} = \text{YAR} - \frac{D_{\text{ds}}}{u} - \left(\frac{v}{s}\right)\max\left(t_r + \frac{u}{2a} - \frac{D_{\text{ds}}}{u}, 0\right) \quad (17)$$

Equation 17 shows how lost time can be reduced by using an upstream detector, a well-known result reported by, for example, Bonneson and McCoy (*7*). With a stopline detector, the critical headway crosses the stopline while the light is green, wasting a lot of time; when the detector is moved upstream, much of the critical headway passes after the signal has transitioned to its clearance interval.

Maximum Green

If the wait for a gap becomes too long, a maximum green constraint will trigger the start of the yellow (the phase "maxes out"). This logic is shown in Figure 2 by the node MaxG positioned a time G_{max} (the maximum green setting) from the start of the green, and a slack arc, whose length must be nonnegative, running from SY to MaxG. When the maximum green is not binding, the slack arc SY–MaxG is harmlessly nonzero, but if the green period's length reaches G_{max}, the nonnegativity constraint on the slack arc will force the start of the yellow.

When the green ends because of max-out, Arcs LXD–GD and LXS–SLA can have any length less than h_{crit}; however, those two arcs will have equal length (h^* in Figure 2), and therefore the position of SLA becomes fixed at $G_{\text{max}} - D_{\text{ds}}/u$.

When there is a max-out, the period of subcritical headways extends from UX all the way to SLA. Accounting for the fraction of that period that is used, and because it may be assumed that the minimum green will not be binding in any cycle that maxes out,

$$L_{\text{gap}}\big|\text{max-out} = 0 \quad (18)$$

$$L_x\big|\text{max-out} = \left(G_{\text{max}} - \frac{D_{\text{ds}}}{u} - t_{\text{EQ}}\right) * \left(1 - \frac{v_{\text{subcrit}}}{s}\right) \quad (19)$$

where $v_{\text{subcrit}} = 1/E[H \mid H < h_{\text{crit}}]$. Equation 19 shows how maximum green constraints improve the efficiency of an actuated signal by eliminating gap lost time. Of course, that efficiency gain is negated if G_{max} is too short to allow the queue to discharge, which prevents G_{max} from being set too aggressively.

In a deterministic approach, like that followed in the HCM, if the expected number of arrivals during an expected cycle results in max-out, max-out will be assumed to always occur, overestimating phase length because max-out will not occur in every cycle; and if expected values do not predict max-out, the maximum green constraint will be assumed nonbinding, again overestimating phase length because maximum green will limit phase length in some cycles.

Instead of the always-or-never approach, the impact of the maximum green on expected gap lost time can be accounted for as follows:

$$L_{\text{gap}} = p_{\text{max-out}} h_{\text{crit}} \quad (20)$$

where $p_{\text{max-out}}$ is the fraction of cycles in which the phase maxes out. Without detailed modeling, it may be difficult to estimate $p_{\text{max-out}}$ with confidence for a given phase; however, one can often make a rough estimate or prior guess of $p_{\text{max-out}}$, in which case Equation 19 models the maximum green effect on a continuous scale rather than as all-or-nothing.

JOINT TERMINATION OF PHASES AT BARRIER

Controllers offer two options to ensure that the two phases that end each half-cycle (at the barrier) end simultaneously. One is simultaneous gap-out, a default setting in at least some American controllers, which requires that both phases must have a gap greater than or equal to minGap at the same moment to force an end to the extension green. In the other option, nonsimultaneous gap-out, normally used in the Netherlands, whichever phase gaps out first enters a green subphase of length G_p called the parallel green until the other phase gaps out (or until max-out), and then both phases end their green.

Nonsimultaneous Gap-Out

An activity network representing nonsimultaneous gap-out of phases i (critical) and j is shown in Figure 3, where maximum green effects have been omitted for clarity. The start of the yellow waits until both phases have independently gapped out, represented by the Event GD_{both}, which follows Nodes GD_i and GD_j with slack arcs representing parallel green time. The parallel green time on the critical phase's slack arc is G^i_{px}.

In the upper part of the diagram, representing traffic flow, LXS_i, is followed by an interval of lost time as the critical gap passes and then by a period of flow with unconstrained headways whose length is G^i_{px} plus the late-arrival period given by Equation 16.

The lost time during the late-arrival period is incorporated into the end lost time formula (Equation 17). Lost time for the parallel green extension, L_{px}, represents the unused time during the parallel green and is given by

$$L_{\text{px}} = \left(1 - \frac{v^i}{s^i}\right) G^i_{\text{px}} \quad (21)$$

When gap-out processes are modeled deterministically, parallel extension lost time will always be zero for critical phases, biasing cycle length estimates downward. The error will be negligible when the demands of the noncritical half-rings are small compared with their parallel critical half-rings but not otherwise.

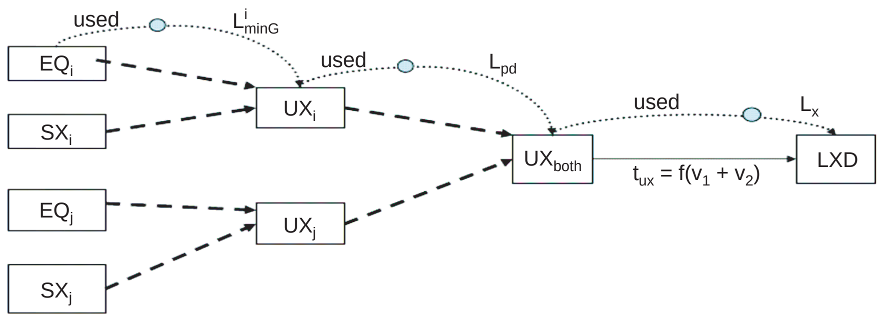

FIGURE 4 Simultaneous gap-out.

Simultaneous Gap-Out

An activity network representing simultaneous gap-out of two phases, *i* (critical) and *j*, is shown in Figure 4. Assuming that minGap is large enough to prevent gap-out if either phase is discharging its queue, the search for a gap does not effectively begin until both phases have discharged their queues and passed their minimum green. This is Event UX_{both}, which follows UX_i and UX_j by slack arcs representing the time during which the first phase that has discharged its queue and passed the minimum green waits for the other to do the same. The fraction of this average slack that is not used is called parallel discharge lost time, L_{pd}, which is the average length of the critical phase's slack arc multiplied by $(1 - v_i/s_i)$.

Following Node UX_{both}, the usual gap-out logic applies, but with one big difference: the headway distribution must represent the combined volumes of both phases, $v_i + v_j$, and should account for the number of lanes serving both phases combined, making the extension green and its associated lost time longer than it would be with nonsimultaneous gap-out, especially on multilane approaches. (It also makes max-out far more likely.) Using volume shares to determine the fraction of the critical phase vehicle represented by those headways, Equation 8 becomes modified as follows:

$$n_{ux} = \left(\frac{p}{(1-p)}\right)\left(\frac{v_i}{(v_i + v_j)}\right) \qquad (8a)$$

with *p* based on the combined (two-directional) headway distribution. Equation 9 for extension lost time is still valid with the headway distribution for phase *i*.

RECALL AND PHASE SKIPPING

If a phase has no queue when its turn in the cycle comes, it will be skipped unless it is set to recall. It is common for left-turn phases, and sometimes minor through phases, not to be set to recall.

The impact of skipping can only be evaluated by using a probabilistic analysis, which is included in the HCM's recommended procedure. The probability of phase *i*'s being skipped assuming a Poisson arrival process is $\exp(-v_i r_i')$, where r_i' is the expected length of phase *i*'s red period minus the intergreen following the conflicting phase preceding phase *i*. (The controller decision of whether to skip is made just before that intergreen phase.)

The skipping of a critical phase increases the chance of a noncritical phase's becoming dominant, increasing parallel lost time and thus reducing some of the apparent benefit of phase skipping. No attempt has been made to estimate this effect.

DIRECT ESTIMATION OF CYCLE LENGTH

Under certain simplifying conditions—nonsimultaneous gap-out, low noncritical demands, no phase skipping, and minimum and maximum green constraints can be ignored—the only lost time components are L_s, L_x, L_{gap}, and L_{end}, of which the first consists of data and the remainder can be determined by formulas (Equations 9, 14, and 17) without prior knowledge of the cycle length. In such a case, average cycle length can be determined directly from Equation 1.

The ability to readily determine lost time components and cycle length makes it easy to explore the impact of demand and design parameters. Figure 5 shows expected cycle length under the simplifying

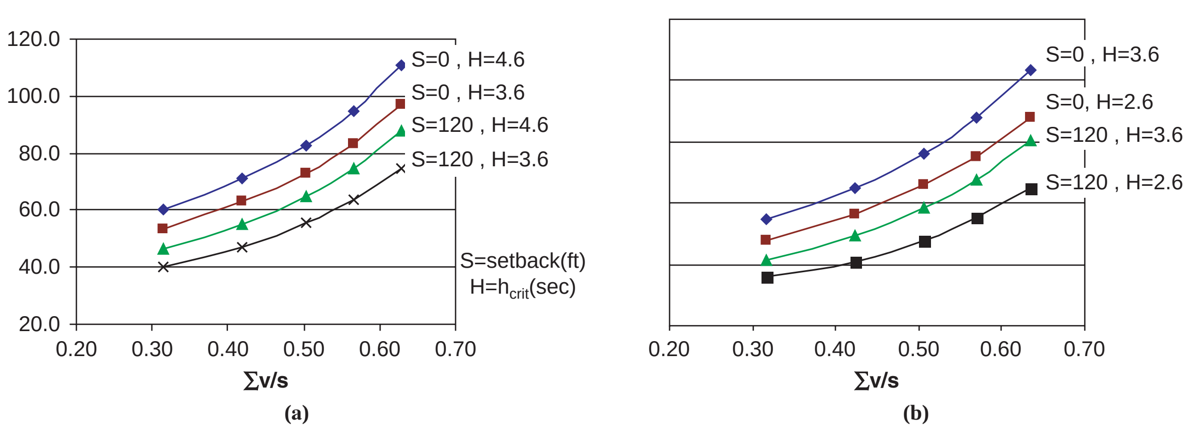

FIGURE 5 Expected cycle length as a function of demand, critical headway *H*, and detector setback *S*: (*a*) one through lane and (*b*) two through lanes.

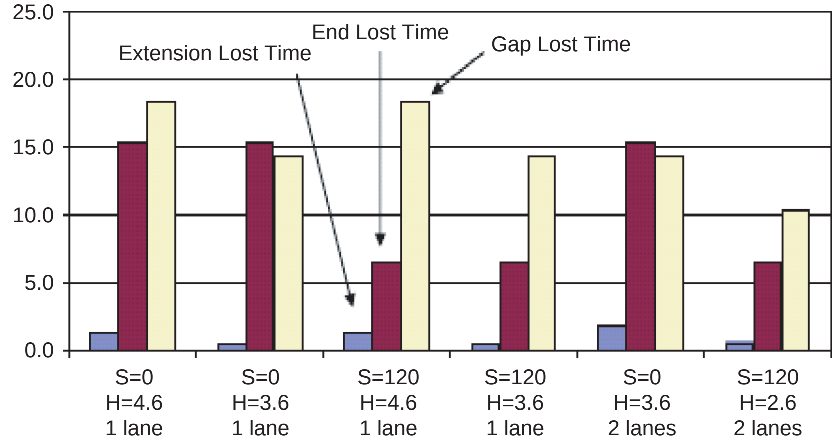

FIGURE 6 Selected lost time components per cycle as a function of critical headway *H* in seconds and detector setback *S* in feet.

conditions mentioned earlier for varying levels of demand ($\Sigma v/s$), detector setback, critical headway, and number of through lanes per approach (number of left-turn lanes per approach is always 1). Critical phase volumes for the single-through-lane case are shown in Figure 1 for low- and high-demand scenarios; intermediate scenarios are linear interpolations. Through volumes are simply doubled to create the two-through-lane scenarios. Saturation flow rate, measured experimentally from the microsimulation model used to verify results, was 1,950 (veh/h)/lane; start-up lost time, also measured experimentally, was 1.5 s for each phase.

Figure 5 shows how sensitive average cycle length is to detector setback and h_{crit}. For a given level of demand, using an upstream detector instead of a stopline detector reduces average cycle length by about 20 s, and making h_{crit} 1 s shorter reduces average cycle length by about 12 s. Reductions of this size make a substantial difference in level of service.

Figure 6 shows how demand and design parameters affect extension, gap, and end lost time for the main street through movement. The main design effect is reducing end lost time by a combination of detector setback and critical gap; also substantial is the effect on gap lost time by reducing h_{crit}. The impact on extension lost time is relatively minor.

VERIFICATION

To verify the model developed, simulation experiments were performed with VISSIM, using its default parameters for vehicle behavior. Each experiment consisted of a 1-h simulation after four cycles of start-up. Figure 7 shows the good agreement in expected cycle length between the model's formula-based results and the simulation experiments. The scenarios reported in Figure 7 include not only those covered in Figure 5 but also the scenarios described later involving minimum green constraints (with recall) and cases with nonsimultaneous gap-out, in which the demand of noncritical phases was as great as 85% of the critical phase demand.

IMPACT OF MINIMUM GREEN CONSTRAINTS

Two sources of randomness affect the distribution of *n*, which is needed to evaluate Equation 5 or 6: random arrivals for a red period of given length *r* and different values of *r* resulting from varying cycle lengths. In most practical situations, the second effect is small compared with the first and may be neglected.

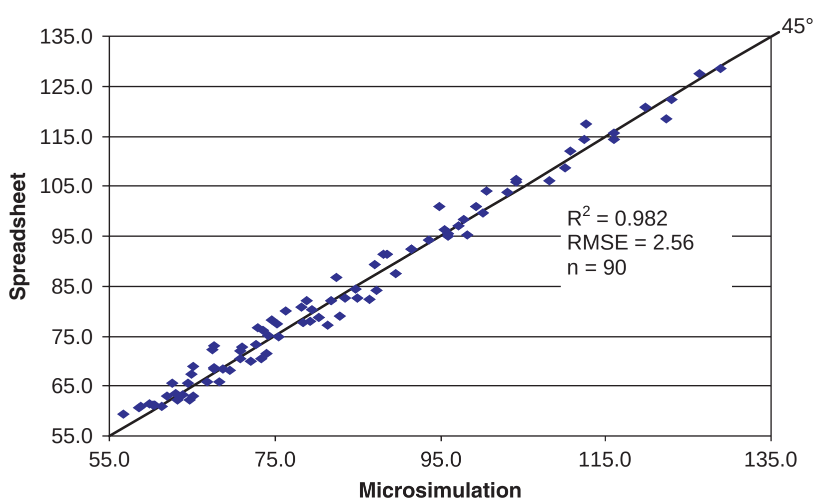

FIGURE 7 Expected cycle length for spreadsheet model versus simulation experiments.

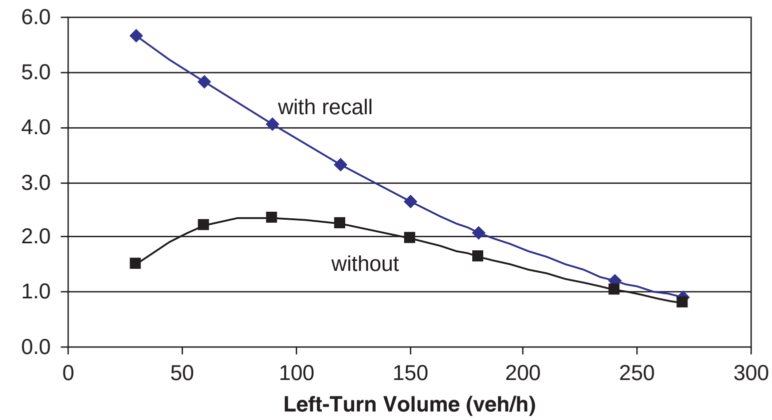

FIGURE 8 Minimum green lost time versus volume with and without recall.

A two-pass procedure amenable to spreadsheet calculation was tested against both iterative spreadsheet calculation and full simulation in the case of 8-s minimum green constraints applied to left-turn movements with and without recall. In the first pass, Equation 1 was used to find cycle length and splits assuming that the left turns were governed by minimum green; in the second pass, n's distribution was Poisson over the phase's red period as determined in the first pass. For each possible value of $n < n_o$, its probability and conditional impact were calculated to evaluate Equation 5 or 6.

The two-pass process yielded cycle length estimates within 0.5 s of average cycle length as determined by using repetitive iterations, confirming the validity of the two-pass process when only minor movements are affected by minimum green constraints and when there is no phase skipping. Errors compared with microsimulation were similarly small.

Figure 8 shows minimum green lost time determined by using simulation results for a single left-turn approach for various levels of demand when demand on other streams is such that the cycle length is about 60 s. (Results when cycle length is about 75 s or 90 s were similar.) When there is no phase skipping, minimum green lost time diminishes with volume, as expected, from a maximum of $(G_{min} - L_s)$. When there is phase skipping, minimum green lost time varies little with demand, holding close to 2 s for a wide range of left-turn volumes.

IMPACT OF NONCRITICAL PHASES AND SIMULTANEOUS GAP-OUT

Until now, all of the results were obtained by assuming that noncritical phases do not affect the cycle length. For the nonsimultaneous gap-out setting, VISSIM and its application programming language VAP were used to determine average parallel extension lost time for a phase terminating at a barrier as a function of the ratio of noncritical to critical phase demand under different levels of overall demand. Results are shown in Figure 9*a*.

For ratios near 1, lost time due to waiting for a parallel phase to gap out was as great as 6 s in the high-demand scenario. However, for ratios between 0.15 and 0.85, L_{px} grew gradually from near 0 to about 2 s and was stable for a wide range of critical phase demands; an equation for predicting L_{px} is shown in Figure 9. As an example, these results indicate that if noncritical demand rises from 40% to 80% of the critical phase demand, the added lost time will be only about 1 s per half-cycle.

The formula for L_{px} was implemented into the spreadsheet model and very good agreement was found between predicted and simulated cycle length; comparisons are included in Figure 7.

Figure 9*b* compares simultaneous and nonsimultaneous gap-out, showing the sum of extension lost time (which for simultaneous gap-out is a function of combined demand in the two directions) and

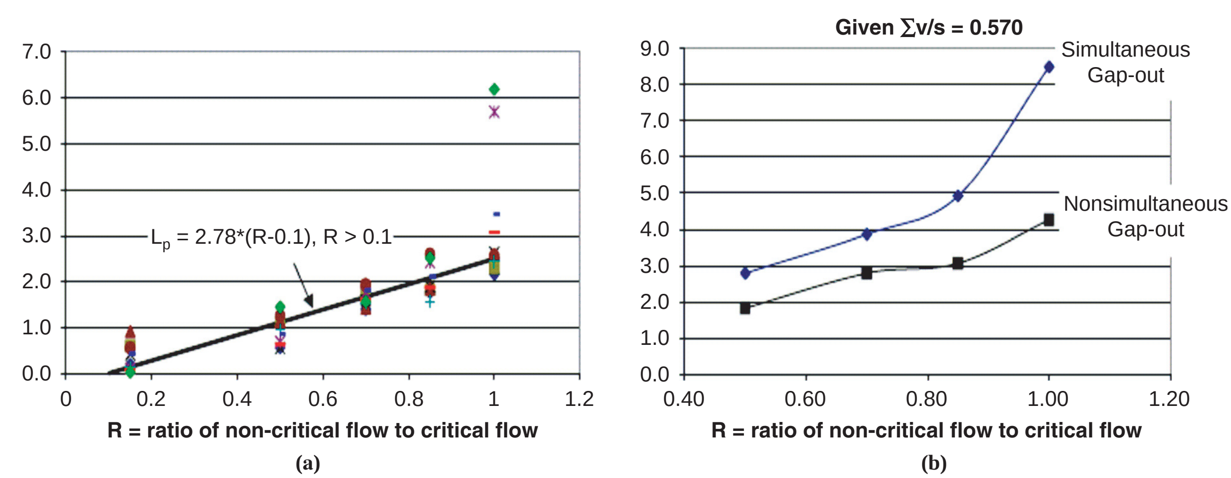

FIGURE 9 Lost time due to termination of green at barrier: (*a*) parallel extension lost time with nonsimultaneous gap-out and (*b*) sum of parallel extension or parallel discharge lost time and extension lost time.

either parallel extension or parallel discharge lost time for a case with moderate demand. Results are from VISSIM simulations, with VAP programming to mark key events. As the graph shows, simultaneous gap-out creates more lost time, especially when the noncritical phase volume is high. A simple formula for estimating L_{pd}. has not been proposed here.

CONCLUSION

Modeling the interactions between signal timing and traffic flow by using an activity network allows one to better understand how actuated traffic signals perform and in particular to understand how demand and design parameters affect average cycle length. The network framework is especially useful for understanding interactions between parallel signal phases in a ring-and-barrier system. It was also shown how deterministic estimation of cycle length, even if iterative, has a downward bias with respect to minimum green and parallel green effects and an upward bias with respect to the maximum green effect.

Seven lost time components were identified, and simple methods for estimating them were proposed for six of them—lost time associated with start-up, minimum green, parallel queue discharge (for simultaneous gap-out), extension green, parallel extension (for nonsimultaneous gap-out), the passing of the critical gap, and phase end. Simulation experiments verify the simple formulas used.

With lost time components and cycle length estimates less of a black box, it is the authors' hope that this approach will allow engineers to focus more clearly on designs that improve actuated signal operation.

REFERENCES

1. Newell, G. F., and E. E. Osuma. Properties of Vehicle-Actuated Signals: II Two-Way Streets. *Transportation Science,* Vol. 3, No. 2, 1969, pp. 99–125.
2. Lin, F.-B. Estimation of Average Phase Durations for Full-Actuated Signals. In *Transportation Research Record 881,* TRB, National Research Council, Washington, D.C., 1982, pp. 65–72.
3. Akcelik, R. Estimation of Green Times and Cycle Length for Vehicle-Actuated Signals. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1457,* Transportation Research Board of the National Academies, Washington, D.C., 1994, pp. 63–72.
4. *Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 2000.
5. Head, L., D. Gettman, D. M. Bullock, and T. Urbanik II. Modeling Traffic Signal Operations with Precedence Graphs. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2035,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 10–18.
6. Head, L., D. Gettman, and Z. Wei. Decision Model for Priority Control of Traffic Signals. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1978,* Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 169–177.
7. Bonneson, J. A., and P. T. McCoy. Methodology for Evaluating Traffic Detector Designs. In *Transportation Research Record 1421,* TRB, National Research Council, Washington, D.C., 1993, pp. 76–81.

The Traffic Signal Systems Committee sponsored publication of this paper.