PERGAMON

Transportation Research Part B 34 (2000) 471–491

TRANSPORTATION
RESEARCH
PART B

www.elsevier.com/locate/trb

# Optimal signal timing for an oversaturated intersection

Tang-Hsien Chang *, Jen-Ting Lin

*Department of Transportation Science, Tamkang University, P.O. Box 7-876, Taipei 10617, Taiwan, ROC*

Received 5 September 1998; accepted 16 June 1999

**Abstract**

Traffic congestion occurs frequently at downtown intersections during rush hours, at road construction zones as well as at accident sites. Under such circumstances, traffic flow exceeds intersection capacity causing queuing of automobiles that cannot be eliminated in one signal cycle. In this paper, we present a timing decision methodology which considers the whole oversaturation period. Discrete dynamic optimization models are developed and an algorithm to solve them is presented. The optimal cycle length and the optimal assigned green time for each approach are determined for the case of two-phase control. The application of the performance index model to certain multi-phase signals in common use is also introduced. Evaluation results indicate that the proposed discrete type performance index model is a more appropriate design for congested traffic signal timing control. 

## 1. Introduction

Multilevel design strategies are a novel trend in traffic signal control (Gartner et al., 1995). Among the promising design features are included the ability to avoid and relieve congestion. Many basic signal theories have been studied in recent decades, including those developed by Webster (1958), May (1965) and Allsop (1972) and that described in the Highway Capacity Manual (1985). However, relatively few of those models have addressed congestion relief strategies. Neither control systems nor commonly used software such as SOAP (1985) and TRANSYT (1987) can adequately handle oversaturated traffic. The performance of these conventional signal systems deteriorates under heavy traffic conditions (Tarnoff and Parsonson, 1981; Cronje, 1983; Elahi et al., 1991). While addressing the limitations of conventional signal control systems, Cronje (1983) developed a model for optimizing fixed-time signalized intersections that can be applied to

* Corresponding author. Tel.: +886-2-2363-1004; fax: +886-2-2622-1135.
*E-mail address:* thchang@im1.im.tku.edu.tw (T.-H. Chang).


PII: S0191-2615(99)00034-X

**Nomenclature**

| | |
|---|---|
| $a_i$ | the number of lanes on approach or movement $i$ |
| $\mathbf{B}$ | the control gain |
| $c$ | cycle length |
| $d$ | the average delay per vehicle in a cycle |
| $D$ | total delay in a cycle |
| $\mathbf{D}(k)$ | total delay in cycle state $k$ |
| $D_i(k+1)$ | delay per lane of approach or movement $i$ in cycle state $k$ |
| $F$ | number of stops |
| $g$ | effective green time |
| $g_{\max}, g_{\min}$ | upper limit and lower limit of effective green time |
| $g(k)$ | green time in cycle state $k$ |
| $g_i(k)$ | effective green time for approach or movement $i$ in cycle state $k$ |
| $g_{i\max}, g_{i\min}$ | upper limit and lower limit of effective green time for approach or movement $i$ |
| $G(k)$ | green-time adjusted factor due to stop penalty in $k$ |
| $G_i(k)$ | green-time adjusted factor due to stop penalty in $k$ for approach or movement $i$ |
| $\mathbf{H}$ | Hamiltonian formula |
| $k$ | the pointer of a cycle state in a sequence of cycles during the saturation period |
| $K$ | stop penalty factor |
| $l_i$ | queue length of approach or movement $i$ |
| $l(k)$ | queue length at the beginning of cycle state $k$ |
| $l_i(k+1)$ | queue length of approach or movement $i$ at the end of cycle state $k$, equal to the queue length of approach or movement $i$ at the beginning of cycle state $k+1$ |
| $N$ | terminative state |
| PI | performance index |
| $\mathbf{PI}(k)$ | total performance index in state $k$ |
| $\mathrm{PI}_i(k)$ | performance index of approach or movement $i$ in state $k$ |
| $q$ | input flow rate |
| $q(k)$ | input flow rate in cycle state $k$ |
| $q_i(k)$ | input flow rate of approach or movement $i$ in cycle state $k$ |
| $s$ | saturated flow rate as defined by Webster (1958) |
| $s_i$ | saturated flow rate of approach or movement $i$ |
| $\mathbf{u}_{\max}, \mathbf{u}_{\min}$ | upper and lower limit of control variables |
| $u(k)$ | control variables in state $k$ |
| $W_i(k)$ | exogenous variable of approach or movement $i$ in state $k$ |
| $\mathbf{W}(k)$ | exogenous variables in state $k$ |
| $Y_i(k)$ | exogenous variable of approach or movement $i$ in state $k$ |
| $\mathbf{Z}(k)$ | exogenous variables in state $k$ |
| $\lambda(k)$ | Lagrange multiplier in state $k$ |
| $\bar{\lambda}$ | $g/c$, ratio of effective green time |
| $\Omega$ | Hessian matrix |

undersaturated and oversaturated conditions. That investigation also compared microscopic and macroscopic models to determine delay and the number of stops, indicating that the macroscopic approach is sufficiently accurate for practical purposes. In a related work, Elahi et al. (1991) developed a knowledge-based system SCII. For near- and over-saturated conditions, SCII adopts the deterministic model proposed by Newell (1982), in which the effects of random variations are neglected since arriving and queuing vehicles provide a steady source of inputs. Elahi et al. (1991) also indicated that the TEXAS model and NETSIM have no optimization capability. Although capable of providing optimal design, SOAP84 heavily depends on Webster's (1958) approach. When searching for timing optimization, the above models only plan for the next single cycle after the executing one, not concurrently for the entire congestion period.

The timing design of isolated signals is a prerequisite for traffic control. This paper presents a novel strategy for the timing decision of isolated signals during congestion or oversaturation. A macroscopic and deterministic model is developed. The underlying notion of the delay formula is derived from the interactive relationship between the delay in a signal cycle and the following cycle. This relationship lacks a conventional signal timing formula. Dispersing a whole queue in one cycle in oversaturated conditions is problematic owing to the maximum cycle length constraint. The remaining automobiles in queues cause delay in each cycle stage, i.e. the delay in a cycle affects the delay in the subsequent cycle. Conventional timing strategies consider only the optimization of a single cycle, commonly referred to as static systems, and are obviously unsatisfactory. Optimal control timing should be designed to regulate the traffic for minimum delay during the entire oversaturated period. A dynamic theory of optimization has to be developed to resolve such a cycle-chaining problem.

Regarding the one-by-one structure of cycle-chaining states, the conventional delay formula can be modified into a state-dependent form called 'state space equations'. An optimal control methodology can then be applied to determine optimal timing from the constructed state space equation of an intersection. Herein, we focus mainly on minimizing total intersection delay during the entire oversaturated period, not per cycle only. The proposed model is formulated as a discrete type operation. Gazis (1964), Gazis and Potts (1965), Green (1968), Burhardt (1971), Kaltenbach and Koivo (1974), Dans and Gazis (1976) and Michalopoulos and Stephanopolos (1977, 1978) constructed similar models for oversaturation control. But their models are all continuous types and do not address the problem of optimizing cycle length. Gazis (1964) proposed that, during an oversaturated period, the queues in all approaches should be allowed to disperse completely and simultaneously, thereby minimizing the total delay (Green, 1968). This method focuses on ensuring that the green time does not have any loss in any cycle during the oversaturated period. This type of control is terminated when completely dispersing the queues of all approaches. Michalopoulos and Stephanopolos (1977, 1978) proposed an efficient two-stage timing method, termed 'bang-bang control', for the controlled signal. Their method attempts to find an optimal switch-over point during the oversaturated period to interchange the timing of the approaches. For example, during the first stage, the procedure is as follows: set maximal green time to the approach having a maximal arrival rate and minimal green time to the minimal arrival rate approach. At the optimal switch-over point, switch the maximal green time to the minimal arrival approach and the minimal green time to the maximal arrival approach.

Continuous type models are limited in that the switch-over point does not necessarily occur at the end of a cycle, neither does the termination of the oversaturated period occur only at the end of the final cycle. On the other hand, the switch-over points determined by a discrete model occur exactly at the termination of a cycle. Discrete operation provides a smooth, regular, and ordered transfer of control. Calculating delay is more reliable. In addition, the penalty level incurred by vehicle stops is easily incorporated. It is more suitable for calculating the optimal cycle length and setting the optimal green time for each approach.

Details of the two-phase model are described below. Certain multi-phase signals in common use are introduced later.

## 2. Two-phase timing plan for oversaturation control

### *2.1. Subjective function with state space representation*

The following describes the two proposed models: one is a basic discrete minimal delay model, and the other a performance index model. The former is to manifest the complexity of the continuous delay model developed by Michalopoulos and Stephanopolos (1977, 1978), and demonstrate that pure delay models are ineffective in searching optimal cycle length. The latter is suggested to be more appropriate in studying oversaturation control.

#### *2.1.1. Discrete minimal delay model*

Fig. 1 illustrates the situation of queue during oversaturation. The graph there represents a queue $l(k+1)$ left when the green time terminates at a certain cycle state $k$. ('Cycle state' denotes the cycle in a sequence of cycles during a saturation period.) The delay in the graph can geometrically be calculated as

$$D = \frac{1}{2}\left[2 \cdot l(k)c + q(k)c^2 - sg^2(k)\right]. \tag{1}$$

This equation meets the May's delay formula (May, 1965)

$$d = \frac{c(1-\bar{\lambda})^2}{2(1-\bar{\lambda}x)} \tag{2}$$

since $\bar{\lambda} = g/c,\ x = qc/sg$, and $D = d \cdot qc$, the area of the graph in Fig. 1. The continuous delay model developed by Michalopoulos and Stephanopolos (1977, 1978) is also consistent with this approach.

In the case of a cross intersection with a two-phase signal control as Fig. 2 illustrates, during oversaturation, the queue and dispersion situation is as indicated in Fig. 3. Without loss of

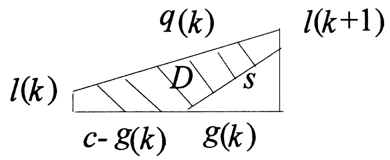


Fig. 1. Situation of the queue in a certain phase during oversaturation.

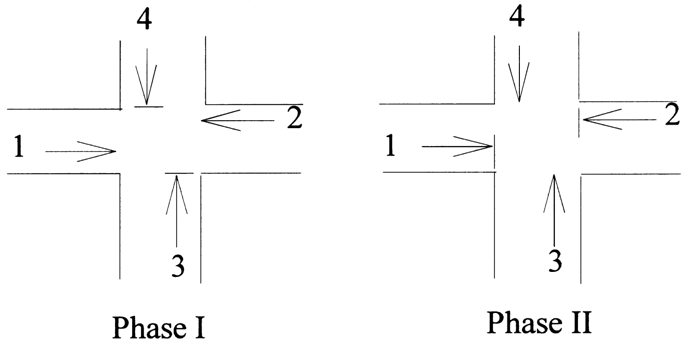


Fig. 2. Four-leg intersection with two-phase signal control.

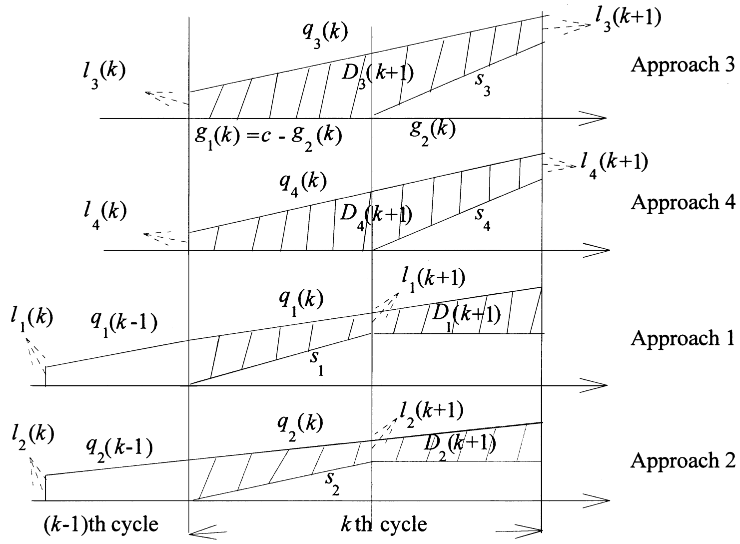


Fig. 3. Queue and delay of a four-leg intersection with two-phase control.

generality, it is assumed herein that the cumulative demand on all the approaches is a linear asymptotic function of time and that the cumulative output curves do not intersect the cumulative input curves for any of the approaches. This fact implies that no queue becomes negative or zero before the end of the oversaturated period. If a queue becomes negative while the signal is green, the designed green time becomes invalid due to the waste of control time.

According to Fig. 3, the relation of the queue lengths between state $k$ and $k+1$ can be represented by the following equations:

$$l_1(k+1) = l_1(k) + q_1(k-1)g_2(k-1) + [q_1(k) - s_1] \cdot [c - g_2(k)], \tag{3a}$$

$$l_2(k+1) = l_2(k) + q_2(k-1)g_2(k-1) + [q_2(k) - s_2] \cdot [c - g_2(k)], \tag{3b}$$

$$l_3(k+1) = l_3(k) + q_3(k)c - s_3 g_2(k), \tag{3c}$$

$$l_4(k+1) = l_4(k) + q_4(k)c - s_4 g_2(k). \tag{3d}$$

Also, from Fig. 3, the delay of approach 1 can be stated as

$$D_1(k+1) = \frac{1}{2}\left[2l_1(k)c + 2q_1(k-1)g_2(k-1)c + q_1(k)c^2 - s_1c^2 + s_1g_2^2(k)\right]. \tag{4}$$

Thus,

$$D_1(k+2) = \frac{1}{2}\left[2l_1(k+1)c + 2q_1(k)g_2(k)c + q_1(k+1)c^2 - s_1c^2 + s_1g_2^2(k+1)\right]. \tag{5}$$

Incorporate (3a) into (5); to obtain

$$D_1(k+2) = D_1(k+1) + \frac{1}{2}s_1g_2^2(k+1) + W_1(k+1) \tag{6}$$

in which,

$$W_1(k+1) = \frac{1}{2}\left[q_1(k)c^2 - s_1g_2^2(k) - 2s_1c^2 + 2s_1g_2(k)c + q_1(k+1)c^2\right]. \tag{7}$$

Similarly, for approach 2, 3, 4 which give

$$D_2(k+2) = D_2(k+1) + \frac{1}{2}s_2g_2^2(k+1) + W_2(k+1), \tag{8}$$

$$D_3(k+2) = D_3(k+1) - \frac{1}{2}s_3g_2^2(k+1) + W_3(k+1), \tag{9}$$

$$D_4(k+2) = D_4(k+1) - \frac{1}{2}s_4g_2^2(k+1) + W_4(k+1), \tag{10}$$

where

$$W_2(k+1) = \frac{1}{2}\left[q_2(k)c^2 - s_2g_2^2(k) - 2s_2c^2 + 2s_2g_2(k)c + q_2(k+1)c^2\right], \tag{11}$$

$$W_3(k+1) = \frac{1}{2}\left[q_3(k)c^2 + s_3g_2^2(k) - 2s_3g_2(k)c + q_3(k+1)c^2\right], \tag{12}$$

$$W_4(k+1) = \frac{1}{2}\left[q_4(k)c^2 + s_4g_2^2(k) - 2s_4g_2(k)c + q_4(k+1)c^2\right]. \tag{13}$$

Suppose that approach 1 has $a_1$ lanes, approach 2 has $a_2$ lanes, approach 3 has $a_3$ lanes, and approach 4 has $a_4$ lanes, the total delay, summed from (6), (8), (9) and (10), should be

$$\begin{aligned} D(k+2) = D(k+1) &+ \frac{1}{2}(a_1s_1 + a_2s_2 - a_3s_3 - a_4s_4)g_2^2(k+1) + a_1W_1(k+1) \\ &+ a_2W_2(k+1) + a_3W_3(k+1) + a_4W_4(k+1). \end{aligned} \tag{14}$$

The above equation can be equivalently restated as a state space expression

$$\mathbf{D}(k+1) = \mathbf{D}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{W}(k), \tag{15}$$

where $\mathbf{D}(k)$ is the state variable; $\mathbf{B}$ the control gain; $\mathbf{u}(k)$ the control variable and $\mathbf{W}(k)$ is the exogenous variable completely known before triggering the state

$$\mathbf{D}(k) = a_1 D_1(k) + a_2 D_2(k) + a_3 D_3(k) + a_4 D_4(k), \tag{16}$$

$$\mathbf{B} = \frac{1}{2}(a_1 s_1 + a_2 s_2 - a_3 s_3 - a_4 s_4), \tag{17}$$

$$\mathbf{u}(k) = g_2^2(k), \tag{18}$$

$$\mathbf{W}(k) = a_1 W_1(k) + a_2 W_2(k) + a_3 W_3(k) + a_4 W_4(k). \tag{19}$$

*2.1.2. Performance index model*

During oversaturation, the number of stops and arrivals will increase. If the stop factor is considered in the utilization of penalty, the model becomes more reasonable (SOAP, 1985; TRANSYT-7F, 1987, 1991). In general, the performance index of signal control is expressed as

$$\mathbf{PI} = D + KF. \tag{20}$$

Applying basic Eq. (20) with (6) and following the derivative procedure of the discrete minimal delay model, we have

$$\mathbf{PI}(k+1) = \mathbf{PI}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{Z}(k), \tag{21}$$

where

$$\mathbf{PI}(k) = a_1 \mathrm{PI}_1(k) + a_2 \mathrm{PI}_2(k) + a_3 \mathrm{PI}_3(k) + a_4 \mathrm{PI}_4(k), \tag{22}$$

$$\mathbf{B} = \frac{1}{2}(a_1 s_1 + a_2 s_2 - a_3 s_3 - a_4 s_4), \tag{23}$$

$$\mathbf{u}(k) = (g_2(k) + G(k))^2, \tag{24}$$

$$\begin{aligned}\mathbf{Z}(k) = &-\frac{1}{2}(a_1 s_1 + a_2 s_2 - a_3 s_3 - a_4 s_4)G^2(k) + a_1 W_1(k) + a_2 W_2(k) + a_3 W_3(k) + a_4 W_4(k) \\ &+ K(a_1 Y_1(k) + a_2 Y_2(k) + a_3 Y_3(k) + a_4 Y_4(k))\end{aligned} \tag{25}$$

satisfying the relationships:

$$\mathrm{PI}_1(k+1) = \mathrm{PI}_1(k) + \frac{1}{2} s_1 g_2^2(k) + K(s_1 - q_1(k))g_2(k) + W_1(k) + KY_1(k), \tag{26}$$

$$\mathrm{PI}_2(k+1) = \mathrm{PI}_2(k) + \frac{1}{2} s_2 g_2^2(k) + K(s_2 - q_2(k))g_2(k) + W_2(k) + KY_2(k), \tag{27}$$

$$\mathrm{PI}_3(k+1) = \mathrm{PI}_3(k) - \frac{1}{2} s_3 g_2^2(k) - K s_3 g_2(k) + W_3(k) + KY_3(k), \tag{28}$$

$$\mathrm{PI}_4(k+1) = \mathrm{PI}_4(k) - \frac{1}{2} s_4 g_2^2(k) - K s_4 g_2(k) + W_4(k) + KY_4(k), \tag{29}$$

$$Y_1(k) = -q_1(k-1)g_1(k-1) + 2q_1(k)c - s_1c, \tag{30}$$

$$Y_2(k) = -q_2(k-1)g_1(k-1) + 2q_2(k)c - s_2c, \tag{31}$$

$$Y_3(k) = -q_3(k-1)c + 2q_3(k)c, \tag{32}$$

$$Y_4(k) = -q_4(k-1)c + q_4(k)c, \tag{33}$$

$$G(k) = \frac{K(a_1s_1 + a_2s_2 - a_3s_3 - a_4s_4 - a_1q_1(k) - a_2q_2(k))}{(a_1s_1 + a_2s_2 - a_3s_3 - a_4s_4)}. \tag{34}$$

### *2.2. Objective function and solution approach*

First, the objective function is set only to minimize the total delay of the entire oversaturated period. The function is assigned to be a quadratic form as below (Anderson and Moore, 1990; Kuo, 1991)

$$\text{MIN } J = \frac{1}{2}(\mathbf{D}(N))^2 + \frac{1}{2}\sum_{k=2}^{N}(\mathbf{D}(k))^2 \tag{35}$$

in which, $N$ is the terminative state of the oversaturated period. Minimizing (35) subjected to (15), based on the optimal control theory, involves equivalently to minimizing the Hamiltonian formula (Kuo, 1991; Luenberger, 1979). The Hamiltonian formula is defined as

$$\mathbf{H} = \frac{1}{2}(\mathbf{D}(k))^2 + \lambda(k+1)[\mathbf{D}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{W}(k)], \tag{36}$$

where the adjoint variables, $\lambda(k),\ k = 0, 1, 2, \ldots$ are functions of time. Now, the task is to find a satisfactory value of the control variable such that (36) is minimal in the subjection of (15). According to the optimal control theory, if an extreme in $\mathbf{H}$ exists, it must satisfy the following conditions:

1. $$\frac{\partial((1/2)\mathbf{D}^2(N))}{\partial\mathbf{D}(N)} = \lambda(N) \Rightarrow \mathbf{D}(N) = \lambda(N), \tag{37}$$

2. $$\frac{\partial\mathbf{H}}{\partial\mathbf{D}(k)} = \lambda(k) \Rightarrow \mathbf{D}(k) + \lambda(k+1) = \lambda(k), \tag{38}$$

3. $$\frac{\partial\mathbf{H}}{\partial\lambda(k+1)} = \mathbf{D}(k+1) \Rightarrow \mathbf{D}(k+1) = \mathbf{D}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{W}(k), \tag{39}$$

4. $$\frac{\partial\mathbf{H}}{\partial\mathbf{u}(k)} = 0. \tag{40}$$

Eq. (36) obviously indicates that only the single control variable $\mathbf{u}(k)$ can minimize $\mathbf{H}$. In signal control, the control variable is related to the green time which should be taken as the value between predetermined upper and lower limits, i.e. $g_{\min} \leqslant g \leqslant g_{\max}$. As defined in (18), $g_2(k)$ is the only control variable here

$$u(k) = \{g_2^2(k) | g_{2\min} \leqslant g_2(k) \leqslant g_{2\max}\}. \tag{41}$$

Eq. (36) clearly reveals that the relation between $\mathbf{H}$ and the control variable $\mathbf{u}(k)$ is linear. This leads to $\partial\mathbf{H}/\partial\mathbf{u}(k) = \lambda(k+1)\mathbf{B} \neq 0$. Thus, to minimize $\mathbf{H}$, the control at each time point should be taken with $\mathbf{u}(k) = g_{2\min}^2$, (i.e. $g_2(k) = g_{2\min}$, $g_1(k) = c - g_{2\min}$) when $\lambda(k+1)\mathbf{B} > 0$; $\mathbf{u}(k) = g_{2\max}^2$, (i.e. $g_2(k) = g_{2\max}$, $g_1(k) = c - g_{2\max}$) when $\lambda(k+1)\mathbf{B} < 0$.

This proposition is clearly a bang-bang control with a two-stage operation, from $g_{2\max}$ switching to $g_{2\min}$ or from $g_{2\min}$ switching to $g_{2\max}$. The time at which the switching is required is termed the 'switch-over' point. For example in Figs. 2 and 3, at the first stage, the maximal green time $g_{2\max}$ is set to approaches 3 and 4 which are associated with higher flow rates; then the green time for the other pair of approaches (with lower flow rates) is determined if the cycle length is given. When the switch-over point evaluated by the described above requirement is reached, the second stage begins, at which point the minimal green time $g_{2\min}$ is switched to substitute $g_{2\max}$ for the control of approaches 3 and 4. The control is terminated at $l_i = 0$ $(i = 1,\ 2,\ 3 \text{ or } 4)$.

If the objective function is to minimize the performance index previously described, the function can also be constructed as (35)–(40), but previous $\mathbf{D}(k)$ should be replaced by $\mathbf{PI}(k)$ and $\mathbf{W}(k)$ be replaced by $\mathbf{Z}(k)$, respectively. This results in the control variable

$$\mathbf{u}(k) = (g_2(k) + G(k))^2 \tag{42}$$

and, which should be operated with the following two situations:

(i) When $\lambda(k+1)\mathbf{B} > 0$, $\mathbf{u}(k) = \mathrm{MIN}(g_2(k) + G(k))^2$. Since only green time $g_2$ can be controlled, $\mathrm{MIN}(g_2(k) + G(k))^2$ implies that $g_2(k) = -G(k)$ if $-G(k)$ is located in the interval $(g_{2\min}, g_{2\max})$; $g_2(k) = g_{2\max}$ if $-G(k) \geqslant g_{2\max}$; and $g_2(k) = g_{2\min}$ if $-G(k) \leqslant g_{2\min}$.

(ii) When $\lambda(k+1)\mathbf{B} < 0$, $\mathbf{u}(k) = \mathrm{MAX}(g_2(k) + G(k))^2$. $\mathbf{u}(k)$ may be verified as a concave curve by taking twice differential of $\mathbf{u}(k)$ with $g_2(k)$. This implies that maximal $\mathbf{u}(k)$, $g_2(k)$ should be at its boundary, $g_2(k) = g_{2\min}$ or $g_2(k) = g_{2\max}$. If $-G(k)$ is located in the interval $(g_{2\min},\ g_{2\max})$, the choice is dependent upon $\mathrm{MAX}\{(g_{2\min} + G(k))^2,\ (g_{2\max} + G(k))^2\}$, i.e. $g_2(k) = g_{2\min}$ when $(g_{2\min} + G(k))^2 > (g_{2\max} + G(k))^2\}$ and $g_2(k) = g_{2\max}$ when $(g_{2\max} + G(k))^2\} > (g_{2\min} + G(k))^2$. In addition, if $-G(k) \geqslant g_{2\max}$, $g_2(k) = g_{2\min}$ is selected. If $-G(k) \leqslant g_{2\min}$, $g_2(k) = g_{2\max}$. Definitely, $g_1(k) = c - g_2(k)$.

Obviously, the control by the performance index model differs somewhat from that by the discrete minimal delay model. In condition (i), when $-G(k)$ drops into the interval $(g_{2\min}, g_{2\max})$, there exists an otherwise condition from 'bang-bang control', i.e. the optimal green time may not be at the assigned boundary. The fact that such an outcome is seldom but still possible during oversaturation (the outcome is much relying on the factor $K$) accounts for why a 'bang-bang like control' is denoted as such a model's control to discriminate the real 'bang-bang control'.

## 3. Algorithm for solving the timing models

Based on the solution approach and the conditions described in Section 2.2, the algorithm for solving the discrete minimal delay timing model is arranged in the following steps:

*Step 1*. Let $k = 2$, and initiate $\lambda(2)$ with a positive value.
*Step 2*. When $\lambda(k)\mathbf{B} > 0$, $g_2(k-1) = g_{2\min}$; when $\lambda(k)\mathbf{B} < 0$, $g_2(k-1) = g_{2\max}$.
*Step 3*. Calculate $\mathbf{D}(k)$, $l_i(k)$ ($i=1, 2, 3, 4$). Check the queue length of each approach, if spillover, adjust $\lambda(2)$ and back to step 1.
*Step 4*. If $l_i(k) < 0$ ($i=1, 2, 3, 4$), go to step 7.
*Step 5*. If $l_1(k) < 0$, $l_2(k) < 0$, $l_3(k) > 0$ and $l_4(k) > 0$, next step; otherwise, employing Eq. (38) calculate $\lambda(k+1)$ and reset $k = k+1$, then go to step 2.
*Step 6*. Let $g_2(k) = g_2(k-1)$. Calculate $l_3(k+1)$ and $l_4(k+1)$. If $l_3(k+1) < 0$ and $l_4(k+1) < 0$, next step; otherwise, adjust $\lambda(2)$ then go to step 1.
*Step 7*. Calculate total delay: sum of $\mathbf{D}(j)$ from $j=2$ to $k-1$. Also, show $g_2(j)$ ($j = 1, \ldots, k-1$); then stop.

Step 1 aims to give priority of dispersion with a maximal green time to the approaches with the maximum flow rate. In step 6, if the calculation falls into 'otherwise', this indicates that there is some green time loss, because the dispersal curve intersects the arrival curve before achieving the termination. Thus, $\lambda(2)$ should be adjusted. Based on Eq. (39), move $\mathbf{D}(k)$ to the right, and substitute $\lambda(k)$ by its elder generations till $k = 2$; then implies

$$\lambda(k) = \lambda(2) - [\mathbf{D}(2) + \mathbf{D}(3) + \cdots + \mathbf{D}(k-1)]. \tag{43}$$

Therefore, in order to make $\lambda(k) < 0$, $\lambda(2)$ should be reduced in the next iteration.

As for the algorithm for solving the performance index model, except for step 2 needing to be replaced by (i) and (ii) described in Section 2.2, the procedure is similar to that for solving the delay timing model.

## 4. Evaluation with a case

Based on the complicated description above, a simplified case is now presented to demonstrate how the model is employed.

### *4.1. Complexity of the continuous delay model*

As mentioned earlier, Michalopoulos and Stephanopolos (1977, 1978) proposed a continuous signal timing model for oversaturated control. The model attempts to minimize the total delay of the entire oversaturated period. Its cycle length is fixed at 150 s ($c = 150$). Also mentioned earlier was its deficiency, i.e. the possible mis-timing of its switch-over point before a cycle is complete. To explain this phenomenon, the case in their paper is applied herein. Assume an intersection of two one-way streets with a two-phase signal control. One street, denoted as approach 1, has two lanes; the other, denoted as approach 2, has a single lane. No left-turn traffic is considered. Approach 1 is with $s_1 = 1400$ pcu/h, $g_{1\max} = 0.65c$, $g_{1\min} = 0.4c$, and approach 2, $s_2 = 1000$ pcu/h, $g_{2\max} = 0.6c$, $g_{2\min} = 0.35c$. Table 1 lists the input volumes (extracted from Michalopoulos and Stephanopolos, 1978). While corresponding to the previously described algorithm for the discrete minimal delay timing model, Table 2 summarizes those results.

Table 1
Five-minute cumulative volumes

| Approach 1 (veh/lane) | Approach 2 (veh/lane) | Time (s) |
|---|---|---|
| 121 | 86 | 300 |
| 205 | 147 | 600 |
| 268 | 192 | 900 |
| 318 | 227 | 1200 |
| 359 | 257 | 1500 |
| 396 | 283 | 1800 |
| 430 | 307 | 2100 |
| 462 | 330 | 2400 |
| 492 | 352 | 2700 |
| 523 | 373 | 3000 |
| 552 | 394 | 3300 |
| 582 | 415 | 3600 |
| 611 | 436 | 3900 |
| 640 | 457 | 4200 |

Table 2
Queue length and control strategy by the discrete minimal delay model

| Cycle sequence ($k$) | Queue length on approach 1 $l(k+1)$ | Queue length on approach 2 $l(k+1)$ | Green time on approach 2 $g_2(k)$ |
|---|---|---|---|
| 1 | 26 | 28 | 52.5 |
| 2 | 48 | 57 | 52.5 |
| 3 | 59 | 73 | 52.5 |
| 4 | 63 | 89 | 52.5 |
| 5 | 60 | 97 | 52.5 |
| 6 | 54 | 105 | 52.5 |
| 7 | 43 | 107 | 52.5 |
| 8 | 38 | 100 | 90.0 |
| 9 | 38 | 90 | 90.0 |
| 10 | 35 | 80 | 90.0 |
| 11 | 32 | 68 | 90.0 |
| 12 | 27 | 56 | 90.0 |
| 13 | 22 | 43 | 90.0 |
| 14 | 15 | 30 | 90.0 |
| 15 | 9 | 16 | 90.0 |
| 16 | 1 | 3 | 90.0 |
| 17 | −7 | 11 | 90.0 |
| Total delay from 1st cycle to 16th | 402,624 (s) | 335,290 (s) | |

According to that table, at the eighth cycle, the control strategy is switched from $g_{2\min}$ to $g_{2\max}$. Notably, the oversaturation control is terminated at the 16th cycle. The total delay summed from approaches 1 and 2 is 737,914 veh-s, which is equivalent to 283.29 s/veh in average delay.

According to Michalopoulos and Stephanopolos (1978), the switch-over point of their continuous delay model is at 994 s, and termination at 2558 s. However, in Table 2, the switch-over point of the discrete delay model is at 1050 s and termination at 2400 s. The switch-over point and the termination of the continuous model are not located at the ends of their corresponding cycles. This means these two cycle lengths are probably not a constant at 150 s. Consequently, the operation would be problematic for a general signal controller. In addition, the oversaturation control time in the continuous model is longer than in the discrete model, indicating that the discrete model is better than the continuous model.

### 4.2. Control of the performance index model

Also, this study applies the case in Section 4.1 to discover the control (bang-bang like control) of the performance index model. The control is the same as the previous strategy for the discrete minimal delay model shown in Table 2. The total delay is extended to 1,046,886 veh-s (equivalent to 363.51 s/veh in average) because the performance index model includes a stop-penalty item.

### 4.3. Comparison of equal time-sharing control and bang-bang like control

At oversaturation, in the bang-bang control, the basic strategy involves assigning the sequence of the maximal green time and minimal green time to a relevant approach. Thus, the traffic of the relevant approach is controlled in two stages: first with maximal green time and then with minimal green time, or vice versa. Regarding the bang-bang like control, results differ somewhat from the bang-bang control, as described in Section 2.2. For nearly all conventional timing designs, the timing split is distributed with the Webster's frame, which maintains the green time ratio with respect to the following formula (Webster, 1958):

$$g_a/g_b = (q_a/s_a)/(q_b/s_b) \tag{44}$$

in which $g_a$, $q_a$ and $s_a$ are the green time volume, and saturation flow rate of approach a, and $g_b$, $q_b$ and $s_b$ the green time, volume, and saturation flow rate of approach b. Obviously, at oversaturation, $g_a : g_b$ should be 1:1. With conventional control, the green time for each approach should be $0.5c$ during oversaturation and equally distributed. Thus, the signal timing becomes equal time-sharing for oversaturation control. The case in Section 4.1 is applied to compare the two strategies. Also investigated herein are the results based on the performance index model and $c = 155$ s involving both the two strategies of equal time-sharing control and bang-bang like control. This study also determines the total delay arising from the conventional equal time-sharing control to be 449.07 s/veh; meanwhile, for bang-bang like control, it is only 362.78 s/veh. This observation confirms Michalopoulos and Stephanopolos' hypothesis on applying bang-bang control (as well as the bang-bang like control proposed herein) to the oversaturation operation to be feasible. Clearly, conventional controls for congested intersections are invalid.

In terms of robustness, bang-bang like control is superior to conventional equal time-sharing control. Providing a ±5% detection or prediction error in traffic volume is acceptable, similar to the above case, we find no difference in the switch-over point when applying bang-bang like control; but, it changes when applying equal time-sharing control. This indicates that bang-bang like control is more stable than equal time-sharing control.

### 4.4. Optimal cycle length

The above discussion confirms that the cycle length remains constant. In fact, the total intersection delay is a function not only of green time $g(k)$, but also of cycle length $c$. For other operational reasons, cycle lengths generally have upper and lower limits. Under normal conditions, the imposed cycle length ranges from 60 to 180 s.

The previous case still applies to the analysis of the optimal cycle length. The left column in Table 3 lists the switch-over points, termination, and average delays with respect to the minimal delay model in each given cycle, varying from 60 to 180 s. This table reveals that the average delay decreases as the cycle length descends. This phenomenon resembles the model's finding for an undersaturated situation, which can be verified from the May's formula presented in Eq. (2). This observation confirms that the delay in $c = 120$ is twice the delay in $c = 60$. However, the descendent slope of the model's finding applied for an undersaturated situation is steeper than in an oversaturated one. Obviously, the optimal cycle length in the minimal delay model should be as small as possible. This finding implies that, at oversaturation, the cycle length is set in $c = 60$. In fact, a short cycle causes more stops (including full stops and partial stops, see TRANSYT-7F User Guide, 1991), resulting in increased operating cost, exhaust energy and pollution. Therefore, pertinent indicates that long cycle lengths are common in practice. Nevertheless, extremely long cycle lengths may also waste time and become unfair. To attain more reasonable control, except

Table 3
Switch-over point, termination and average delay in each given cycle

| Cycle length (s) | Minimal delay model | | | Performance index model | | |
|---|---|---|---|---|---|---|
| | Switch-over point (s) | Termination (s) | Average delay (s/veh) | Switch-over point (s) | Termination (s) | Average delay (s/veh) |
| 180 | 1080 | 2340 | 292.81 | 1080 | 2340 | 365.40 |
| 175 | 1050 | 2275 | 295.71 | 1050 | 2275 | 370.29 |
| 170 | 1020 | 2380 | 294.70 | 1020 | 2380 | 370.34 |
| 165 | 990 | 2310 | 299.39 | 990 | 2310 | 377.60 |
| 160 | 1120 | 2240 | 284.45 | – | – | – |
| 155 | 1085 | 2325 | 284.04 | 1085 | 2325 | 362.78 |
| 150 | 1050 | 2400 | 283.29 | 1050 | 2400 | 363.51 |
| 145 | 1015 | 2320 | 288.53 | 1015 | 2320 | 371.96 |
| 140 | 980 | 2380 | 288.30 | 980 | 2380 | 373.65 |
| 130 | 1040 | 2340 | 280.14 | 1040 | 2340 | 368.47 |
| 125 | 1000 | 2375 | 279.72 | 1000 | 2375 | 370.40 |
| 115 | 1035 | 2300 | 274.44 | 1035 | 2300 | 369.50 |
| 110 | 990 | 2310 | 277.92 | 990 | 2310 | 377.40 |
| 105 | 1050 | 2310 | 268.33 | 1050 | 2310 | 386.18 |
| 100 | 1000 | 2400 | 268.88 | 1000 | 2400 | 373.02 |
| 90 | 990 | 2340 | 268.56 | 990 | 2340 | 381.66 |
| 80 | 960 | 2320 | 269.40 | 960 | 2320 | 394.00 |
| 75 | 975 | 2325 | 265.43 | 975 | 2325 | 395.20 |
| 70 | 980 | 2380 | 259.58 | 980 | 2380 | 394.28 |
| 65 | 975 | 2340 | 260.76 | 975 | 2340 | 404.65 |
| 60 | 960 | 2340 | 260.42 | 960 | 2340 | 414.15 |

for the delay term, considering appropriate stops penalty into the timing determination to reduce relative stops is acceptable. This perspective suggests that the performance index model is more appropriate than the minimal delay model for congestion control. Searching for the optimal cycle length in the discrete performance index model is much more complicated than by a trial and error method. The right column of Table 3 summarizes the searching results while assuming that $K = 30$. On this occasion, the optimal cycle length is 155 s. It is not optimal for the cycle length to become small. This finding also reflects the fact that Michalopoulos and Stephanopolos' (1977, 1978) continuous delay model has difficulty in searching for an optimal cycle length for oversaturation control since those investigators failed to consider stops penalty into their model.

Previous discussion of bang-bang like control did not consider the variation of the lengths of maximal and minimal green time. Maximal and minimal green time are normally regulated by laws in many countries. The regulation can be considered only as constraints. From the mathematical process depicted in previous sections or a theoretical perspective, maximal and minimal green times are actually a boundary condition. An attempt is currently underway to change the boundary values before determining what one cycle length for the optimal control is. Though, the assigned boundary must be within legality. Table 4 lists the details of analysis. In this case, we can conclude that the final optimal control should be set in $C = 160$ s, $g_{2\,\min} = 0.25c$ and $g_{2\,\max} = 0.65c$. The average delay is 320.43 s. Table 4 also indicates that not all provided cycle length can adhere to the warrant of simultaneous dispersion. In other words, not all provided cycle lengths can be applied to oversaturation control.

## 5. Multi-phase timing plans

The above discussion confirms that the discrete type performance index model is quite appropriate for traffic signal timing design. The following introduces the performance index model as applied in commonly employed multi-phase signals.

Herein, two common types of multi-phases control for a four-leg intersection are presented. Fig. 4 displays a three-phase signal arrangement for a left-turn traffic protection. Fig. 5 illustrates the queue and dispersion situation. Based on Fig. 5 and following the process presented in Section 2.1, it yields the form

$$\mathrm{PI}(k+2) = \mathrm{PI}(k+1) + f(g_1, g_3, k+1). \tag{45}$$

Table 5 provides the detail formulae. Also, Fig. 6 illustrates a four-phase signal with left-turn protection and Fig. 7, the queue and dispersion situation. From Fig. 7, we can infer that

$$\mathrm{PI}(k+2) = \mathrm{PI}(k+1) + f(g_1, g_3, g_4, k+1). \tag{46}$$

Table 6 presents the detail formulae of the four-phase model. Solving the multi-phase models for optimal control timing resembles the description for the two-phase model in Section 3. Nevertheless, the control variables in the multi-phase models are not single. Three-phase signals have two control variables and four-phase signals have three. Such a provision leads to a multiple-variables-single-object problem. The decision procedure is quite complicated. In the following, we only briefly provide the solving algorithm of the four-phase model. Eq. (46) can be rewritten as Eq. (21), but

Table 4
Relation between average delay and cycle length under different green time[a]

| Cycle length | $g_{2\,min}$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.25*c* | | | 0.30*c* | | | 0.35*c* | | | 0.4*c* | | |
| | $g_{2\,max}$ | | | | | | | | | | | |
| | 0.55*c* | 0.6*c* | 0.65*c* | 0.55*c* | 0.6*c* | 0.65*c* | 0.55*c* | 0.6*c* | 0.65*c* | 0.55*c* | 0.6*c* | 0.65*c* |
| 180 | **375.70** | 360.73* | – | 396.24 | 355.14 | 341.01* | 393.38 | 365.40 | 355.37* | 406.43 | 389.3* | 382.87* |
| 175 | 376.43* | 365.84* | – | 400.04* | 360.39* | 346.77* | 397.32 | 370.29* | 360.68* | 394.14* | 377.97* | 371.24 |
| 170 | 375.76 | – | – | 398.81 | 361.92 | – | 395.63 | 370.34 | – | **391.39** | **376.89** | 372.83 |
| 165 | 382.05 | – | – | **372.50** | 368.77* | – | 401.94* | 377.60* | **345.57** | 397.86 | 383.20 | 377.60* |
| 160 | 381.04 | 341.98* | **320.43** | 379.37* | 375.80* | 333.63* | 400.51* | 361.77* | – | 408.20 | 382.41 | 384.23* |
| 155 | 388.95 | **344.49** | 330.86* | 379.32 | 378.53* | **337.08** | 384.47 | **362.78** | 346.54 | 402.97 | 390.25* | 370.73 |
| 150 | 388.45 | 346.42 | – | 378.53 | 353.94* | – | **383.16** | 363.51 | 356.44* | 401.49 | 385.16* | 371.20 |
| 145 | 395.88* | 355.71* | – | 386.34 | 357.46 | – | 390.91 | 371.96 | – | 394.27 | 382.95 | 379.33* |
| 140 | 395.99 | 358.60 | 316.98* | 386.23 | 359.71 | 329.38* | 390.50 | 373.65 | – | 391.45 | 383.36 | **366.23** |
| 130 | 409.14 | – | 333.85* | 399.71 | 377.00* | 340.89* | 403.51 | 368.47 | 347.52 | 405.43 | 384.07 | 382.17* |
| 125 | 406.12* | 344.65 | – | 399.90 | **351.80** | – | 403.41 | 370.40 | – | 404.53 | 384.86 | 370.08 |

[a] *Note*: '*' expresses that at termination a queue still exists in a certain approach.'–' expresses that the termination fails in the given condition.

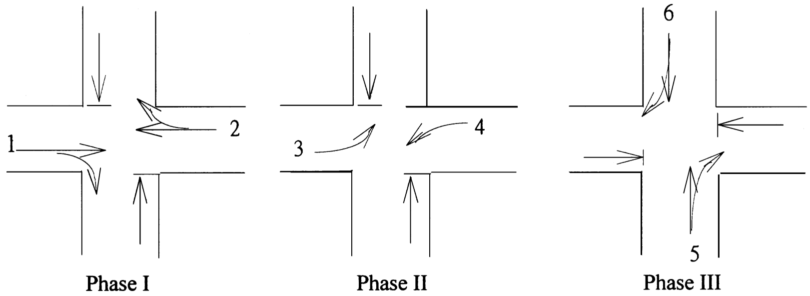


Fig. 4. Three-phase signal with left-turn protection.

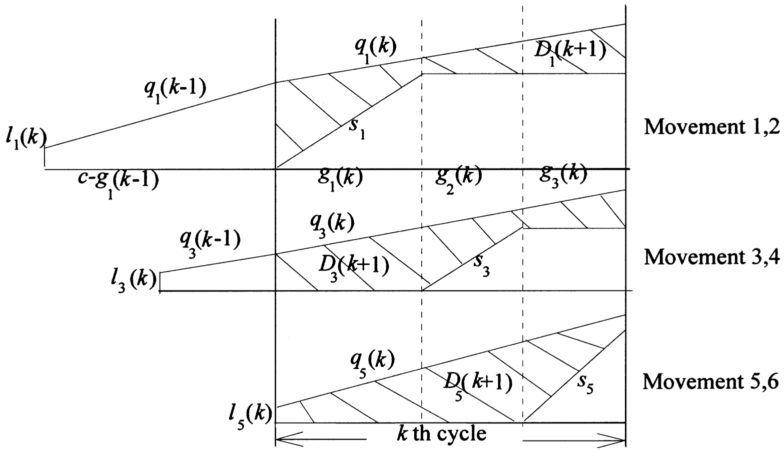


Fig. 5. Queue and delay of a three-phase signal with left-turn protection.

$$\mathbf{B} = \frac{1}{2}\begin{bmatrix} (a_1s_1 + a_2s_2 - a_3s_3 - a_4s_4 - a_5s_5 - a_6s_6) \\ (a_3s_3 + a_4s_4) \\ (a_5s_5 + a_6s_6 - a_7s_7 - a_8s_8) \end{bmatrix}^{\mathrm{T}}, \tag{47}$$

$$\mathbf{u}(k) = \begin{bmatrix} (g_1(k) + G_1(k))^2 \\ (g_3(k) + g_4(k) + G_{34}(k))^2 \\ (g_4(k) + G_4(k))^2 \end{bmatrix}. \tag{48}$$

The objective function attempts to minimize the performance index and the Hamiltonian formula are denoted as Section 2. Since Eq. (48) includes three implicit variables of green time $g_1$, $g_3$ and $g_4$, to find minimal $\mathbf{H}$ from $\partial\mathbf{H}/\partial\mathbf{u}(k) = \mathbf{0}$, a Hessian matrix $\Omega$ should be checked.

$$\Omega = \begin{bmatrix} \partial^2\mathbf{H}/(\partial g_1)^2 & \partial^2\mathbf{H}/\partial g_1\partial g_3 & \partial^2\mathbf{H}/\partial g_1\partial g_4 \\ \partial^2\mathbf{H}/\partial g_1\partial g_3 & \partial^2\mathbf{H}/(\partial g_3)^2 & \partial^2\mathbf{H}/\partial g_3\partial g_4 \\ \partial^2\mathbf{H}/\partial g_1\partial g_4 & \partial^2\mathbf{H}/\partial g_3\partial g_4 & \partial^2\mathbf{H}/(\partial g_4)^2 \end{bmatrix} = \begin{bmatrix} \lambda\cdot\mathbf{b}_1 & 0 & 0 \\ 0 & \lambda\cdot\mathbf{b}_2 & \lambda\cdot\mathbf{b}_2 \\ 0 & \lambda\cdot\mathbf{b}_2 & \lambda\cdot(\mathbf{b}_2 + \mathbf{b}_3) \end{bmatrix}, \tag{49}$$

where $\mathbf{b}_i$ denotes the three elements of vector $\mathbf{B}$ of Eq. (47).

Table 5
A typical three-phase model

$$\mathrm{PI}(k+2) = \mathrm{PI}(k+1) + f(g_1, g_3, k+1)$$

$$f(g_1, g_3, k+1) = \frac{1}{2}(a_1s_1 + a_2s_2 - a_3s_3 - a_4s_4)(g_1(k+1) + G_1(k+1))^2 + \frac{1}{2}(a_3s_3 + a_4s_4 - a_5s_5 - a_6s_6)(g_3(k+1) + G_3(k+1))^2 + Z(k+1)$$

$$G_1(k+1) = -c - K + \frac{K(a_1q_1(k+1) + a_2q_2(k+1))}{(a_1s_1 + a_2s_2 - a_3s_3 - a_4s_4)}$$

$$G_3(k+1) = K - \frac{K(a_3q_3(k+1) + a_4q_4(k+1))}{(a_3s_3 + a_4s_4 - a_5s_5 - a_6s_6)}$$

$$Z(k+1) = -\frac{1}{2}(a_1s_1 + a_2s_2 - a_3s_3 - a_4s_4)(G_1^2(k+1) - c^2) - \frac{1}{2}(a_3s_3 + a_4s_4 - a_5s_5 - a_6s_6)G_3^2(k+1) + \sum_{i=1}^{6} a_i W_i(k+1) + K\sum_{i=1}^{6} a_i Y_i(k+1)$$

$$W_1(k+1) = \frac{1}{2}\left[q_1(k)c^2 - s_1g_1^2(k) - s_1c^2 + q_1(k+1)c^2\right]$$

$$W_2(k+1) = \frac{1}{2}\left[q_2(k)c^2 - s_2g_1^2(k) - s_2c^2 + q_2(k+1)c^2\right]$$

$$W_3(k+1) = \frac{1}{2}\left[s_3(c - g_1(k))^2 - s_3g_3^2(k) + q_3(k)c^2 - 2s_3g_2(k)c + q_3(k+1)c^2\right]$$

$$W_4(k+1) = \frac{1}{2}\left[s_4(c - g_1(k))^2 - s_4g_3^2(k) + q_4(k)c^2 - 2s_4g_2(k)c + q_4(k+1)c^2\right]$$

$$W_5(k+1) = \frac{1}{2}\left[q_5(k)c^2 + s_5g_3^2(k) - 2s_5g_3(k)c + q_5(k+1)c^2\right]$$

$$W_6(k+1) = \frac{1}{2}\left[q_6(k)c^2 + s_6g_3^2(k) - 2s_6g_3(k)c + q_6(k+1)c^2\right]$$

$$Y_1(k+1) = q_1(k+1)c - q_1(k)g_1(k)$$

$$Y_2(k+1) = q_2(k+1)c - q_2(k)g_1(k)$$

$$Y_3(k+1) = -s_3c - q_3(k)c + 2q_3(k+1)c + q_3(k)g_3(k)$$

$$Y_4(k+1) = -s_4c - q_4(k)c + 2q_4(k+1)c + q_4(k)g_3(k)$$

$$Y_5(k+1) = -q_5(k)c + 2q_5(k+1)c$$

$$Y_6(k+1) = -q_6(k)c + 2q_6(k+1)c$$

(i) When $\Omega$ is positive definite, minimal $\mathbf{u}(k)$ is required. Since $\mathbf{u}(k) \geqslant \mathbf{0}$, let $\mathbf{u}(k) = \mathbf{0}$ and then we have $g_1(k) = -G_1(k)$, $g_3(k) = G_4(k) - G_{34}(k)$, $g_4(k) = -G_4(k)$ and $g_2 = c - g_1 - g_3 - g_4 - G_1(k) \epsilon(g_{1\min},\ g_{1\max})$, $(G_4(k) - G_{34}(k))\epsilon(g_{3\min},\ g_{3\max})$, and $-G_4(k)\epsilon(g_{4\min},\ g_{4\max})$. If $-G_1(k) \geqslant g_{1\max}$, $g_1(k) = g_{1\max}$; if $-G_1(k) \leqslant g_{1\min}$, $g_1(k) = g_{1\min}$; if $(G_4(k) - G_{34}(k)) \geqslant g_{3\max}$, $g_3(k) = g_{3\max}, \ldots$ and so on.

(ii) When $\Omega$ is negative definite, maximal $\mathbf{u}(k)$ is taken. Since $\mathbf{u}(k)$ is concave, this implies all $g_j(k)$, $j = 1, 3, 4$, should be at their boundaries, $g_j(k) = g_{j\min}$ or $g_j(k) = g_{j\max}$. If $-G_1(k)\epsilon(g_{1\min}, g_{1\max})$, $g_1(k)$ is chosen from either $g_{1\min}$ or $g_{1\max}$ such that $\mathrm{MAX}\{(g_{1\min} + G_1(k))^2,\ (g_{1\max_{G_1}}(k))^2\}$. If $-G_1(k) \geqslant g_{1\max}$, $g_1(k) = g_{1\min}$ is selected; if $-G_1(k) \leqslant g_{1\min}$, $g_1(k) = g_{1\max}$. The choice for $g_4$ is

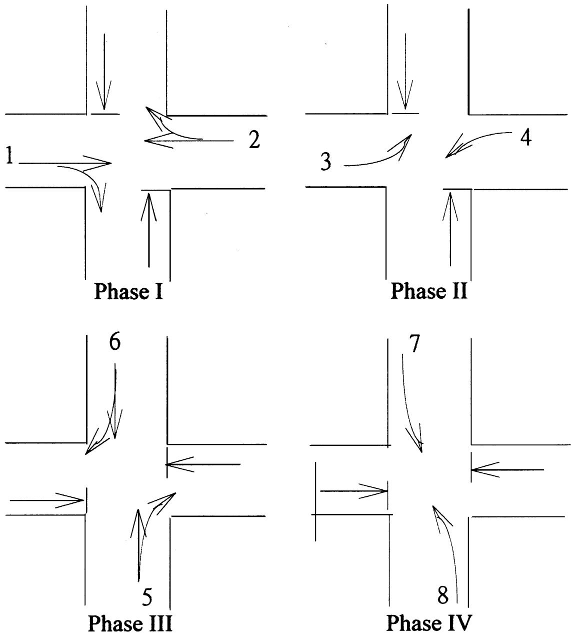


Fig. 6. A four-phase signal with left-turn protection.

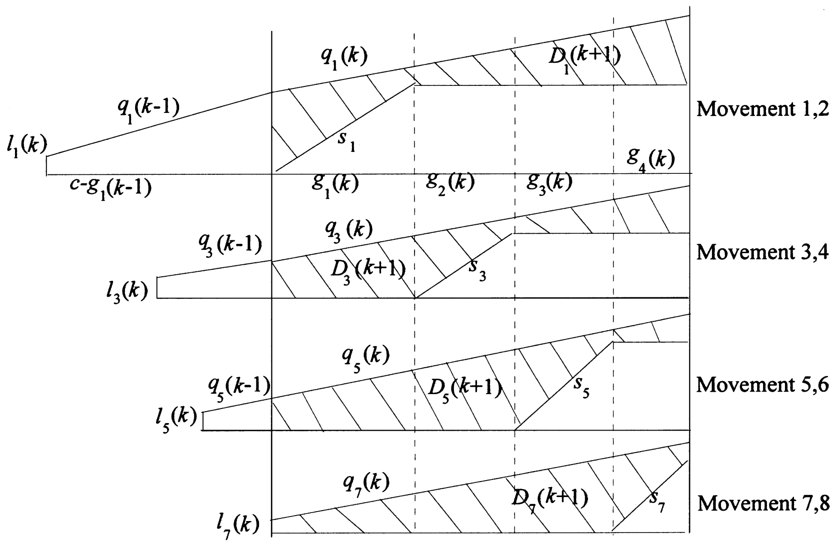


Fig. 7. Queue and delay of a four-phase signal with left-turn protection.

the same as choosing $g_1(k)$. Finally, if $(G_4(k) - G_{34})\epsilon(g_{3\min}, g_{3\max})$, $g_3(k)$ is chosen from which $g_{3\min}$ or $g_{3\max}$ such that $\mathrm{MAX}\{(g_{3\min} + g_4 + G_{34}(k))^2, (g_{3\max} + g_4 + G_{34}(k))^2\}$. If $G_4(k) - G_{34}(k) \geqslant g_{3\max}$, set $g_3(k) = g_{3\min}$; if $G_4(k) - G_{34}(k) \leqslant g_{3\min}$, set $g_3(k) = g_{3\max}$. Certainly, $g_2(k) = c - g_1(k) - g_3(k) - g_4(k)$. The interval $(g_{\min}, g_{\max})$ for left-turn phases is generally shorter than that for through traffic. The cycle set for four-phase signals is longer than for two-phase signals.

Table 6
A typical four-phase model

$$\mathrm{PI}(k+2) = \mathrm{PI}(k+1) + f(g_1, g_3, g_4, k+1)$$

$$f(g_1, g_3, g_4, k+1) = \frac{1}{2}(a_1 s_1 + a_2 s_2 - a_3 s_3 - a_4 s_4 - a_5 s_5 - a_6 s_6)(g_1(k+1) + G_1(k+1))^2 + \frac{1}{2}(a_3 s_3 + a_4 s_4)(g_3(k+1) + g_4(k+1) + G_{34}(k+1))^2 + \frac{1}{2}(a_5 s_5 + a_6 s_6 - a_7 s_7 - a_8 s_8)(g_4(k+1) + G_4(k+1))^2 + Z(k+1)$$

$$G_1(k+1) = -c - K + \frac{K(a_1 q_1(k+1) + a_2 q_2(k+1))}{(a_1 s_1 + a_2 s_2 - a_3 s_3 - a_4 s_4 - a_5 s_5 - a_6 s_6)}$$

$$G_{34}(k+1) = K - \frac{K(a_3 q_3(k+1) + a_4 q_4(k+1))}{(a_3 s_3 + a_4 s_4)}$$

$$G_4(k+1) = K - \frac{K(a_5 q_5(k+1) + a_6 q_6(k+1))}{(a_5 s_5 + a_6 s_6 - a_7 s_7 - a_8 s_8)}$$

$$Z(k+1) = -\frac{1}{2}(a_1 s_1 + a_2 s_2 - a_3 s_3 - a_4 s_4 - a_5 s_5 - a_6 s_6)(G_1^2(k+1) - c^2) - \frac{1}{2}(a_3 s_3 + a_4 s_4)G_{34}^2(k+1) - \frac{1}{2}(a_5 s_5 + a_6 s_6 - a_7 s_7 - a_8 s_8)G_4^2(k+1) + \sum_{i=1}^{8} a_i W_i(k+1) + K\sum_{i=1}^{8} a_i Y_i(k+1)$$

$$W_1(k+1) = 0.5\left[q_1(k)c^2 - s_1 g_1^2(k) - s_1 c^2 + q_1(k+1)c^2\right]$$

$$W_2(k+1) = 0.5\left[q_2(k)c^2 - s_2 g_1^2(k) - s_2 c^2 + q_2(k+1)c^2\right]$$

$$W_3(k+1) = 0.5\left[s_3(c - g_1(k))^2 - s_3(g_3(k) + g_4(k))^2 + q_3(k)c^2 - 2s_3 g_2(k)c + q_3(k+1)c^2\right]$$

$$W_4(k+1) = 0.5\left[s_4(c - g_1(k))^2 - s_4(g_3(k) + g_4(k))^2 + q_4(k)c^2 - 2s_4 g_2(k)c + q_4(k+1)c^2\right]$$

$$W_5(k+1) = 0.5\left[s_5(c - g_1(k))^2 - s_5 g_4^2(k) + q_5(k)c^2 - 2s_5 g_3(k)c + q_5(k+1)c^2\right]$$

$$W_6(k+1) = 0.5\left[s_6(c - g_1(k))^2 - s_6 g_4^2(k) + q_6(k)c^2 - 2s_6 g_3(k)c + q_6(k+1)c^2\right]$$

$$W_7(k+1) = 0.5\left[q_7(k)c^2 + s_7 g_4^2(k) - 2s_7 g_4(k)c + q_7(k+1)c^2\right]$$

$$W_8(k+1) = 0.5\left[q_8(k)c^2 + s_8 g_4^2(k) - 2s_8 g_4(k)c + q_8(k+1)c^2\right]$$

$$Y_1(k+1) = q_1(k+1)c - q_1(k)g_1(k)$$

$$Y_2(k+1) = q_2(k+1)c - q_2(k)g_1(k)$$

$$Y_3(k+1) = -s_3 c - q_3(k)c + 2q_3(k+1)c + q_3(k)(g_3(k) + g_4(k))$$

$$Y_4(k+1) = -s_4 c - q_4(k)c + 2q_4(k+1)c + q_4(k)(g_3(k) + g_4(k))$$

$$Y_5(k+1) = -s_5 c - q_5(k)c + 2q_5(k+1)c + q_5(k)g_4(k))$$

$$Y_6(k+1) = -s_6 c - q_6(k)c + 2q_6(k+1)c + q_6(k)g_4(k))$$

$$Y_7(k+1) = -q_7(k)c + 2q_7(k+1)c$$

$$Y_8(k+1) = -q_8(k)c + 2q_8(k+1)c$$

Otherwise, $\det\Omega = 0$ implies that the control vector is out of control. At this occasion, Eq. (46) should be changed to the alternative forms $f(g_2, g_3, g_4, k+1)$ or $f(g_1, g_2, g_3, k+1)$ or $f(g_1, g_2, g_4, k+1)$; then try to solve it again. When searching $\lambda(k)$ and tracing the optimal solution within the entire oversaturated period, the proposed process described in Section 3 is referred.

## 6. Conclusion

Traffic congestion occurs frequently at downtown signalized intersections during rush hours, in commercial and industrial work zones as well as at traffic incident sites. On such occasions, traffic flow exceeds intersection capacity, causing queuing of automobiles that cannot be eliminated in one signal cycle. According to our results, conventional signal control strategies are inadequate since the designed optimization timing is only considered for the next single cycle after the executing one, not concurrently for the entire congestion period. This paper presents a novel timing decision methodology which takes account of the whole oversaturated period. Two models are derived: one of basic discrete minimal delay model and the other a performance index model. Based on the results presented herein, we can conclude the following:

1. The performance index model is a more appropriate design for congested traffic signal timing control and is superior to the pure minimal delay model;
2. Discrete type models are more applicable in practice than continuous models. Continuous models may activate a switch-over point not occurring at the end of a cycle, whereas the switch-over point in discrete models coincides exactly with that juncture;
3. While corresponding to the findings of previous studies of signal timing for oversaturated intersections, bang-bang like control, by which signals are operated alternatively and sequentially, with minimal maximal green time, significantly outperforms conventional equal time-sharing dispersion control; and
4. Not all provided cycle lengths are applicable to oversaturation control, since some may fail to meet the warrant of simultaneous dispersion indicated by Gazis (1964).

The discrete type performance index model proposed herein, which results in bang-bang like control, is quite appropriate for oversaturation control. The performance of this model is rather robust even when the input data appear to be slightly biased. The proposed model can also determine the optimal cycle length and the optimal assigned green time.

## References

TRANSYT-7F, 1991. User Guide, Methodology for Optimizing Signal Timing, vol. 4. Transportation Research Center, University of Florida.

TRANSYT-7F, 1987. User's Manual, Release 5.0. Federal Highway Administration.

Allsop, R.E., 1972. Delay at a fixed time traffic signal I: theoretical analysis. Transportation Science 6 (3), 260–285.

Anderson, B.D.O., Moore, J.B., 1990. Optimal Control-Linear Quadratic Methods. Prentice-Hall, Englewood Cliffs, NJ.

Burhardt, K.K., 1971. Urban traffic system optimization. Ph.D. Dissertation, University of Minnesota, Minneapolis, Minnesota.

Cronje, W.B., 1983. Optimization model for isolated signalized traffic intersections. Transportation Research Record 905, 80–83.

Dans, G., Gazis, D.C., 1976. Optimal control of oversaturated store-and-forward transportation networks. Transportation Science 10 (1), 1–19.

Elahi, S.M., Radwan, A.E., Goul, K.M., 1991. Knowledge-based system for adaptive traffic signal control. Transportation Research Record 1324, 115–122.

Gartner, N.H., Stamatiadis, C., Tarnoff, P.J., 1995. Development of advanced traffic signal control strategies for intelligent transportation systems: Multilevel design. Transportation Research Record 1494, 98–105.

Gazis, D.C., 1964. Optimal control of a system of oversaturated intersections. Operations Research 12, 815–831.

Gazis, D.C., Potts, R.B., 1965. The oversaturated intersection. In: Proceedings of the Second International Symposium on the Theory of Road Traffic Flow. Organization for Economic Cooperation and Development, Paris, pp. 221–237.

Green, D.H., 1968. Control of oversaturated intersections. Operational Research Quarterly 18 (2), 161–173.

1985. Highway capacity manual. Special Report 209. Transportation Research Board, National Research Council.

Kaltenbach, M., Koivo, H.N., 1974. Modelling and control of urban traffic flow. In: Proceedings of the Joint Automatic Control Conference. University of Texas, Houston, Texas, pp. 147–154.

Kuo, B.C., 1991. Automatic Control Systems, 6th ed. Prentice-Hall, Englewood Cliffs, NJ.

Luenberger, D.G., 1979. Introduction to Dynamic Systems: Theory, Models, & Applications. Wiley, Chichester.

May, A.D., Jr., 1965. traffic flow theory – the traffic engineer's challenge. In: Proceedings of the Institute Traffic Engineering. pp. 290–303.

Michalopoulos, P.G., Stephanopolos, G. Oversaturated signal system with queue length constraints-I. Transportation Research 11, 413–421.

Michalopoulos, P.G., Stephanopolos, G., May 1978. Optimal control of oversaturated intersections theoretical and practical considerations. In: Traffic Engineering & Control. pp. 216–221.

Newell, G.F. 1982. Applications of Queuing Theory, 2nd ed. Chapman & Hall, London, England, pp. 287–300.

SOAP 84, 1985. User's Manual. Federal Highway Administration.

Tarnoff, P.J., Parsonson, P.S., 1981. Selecting traffic signal control at individual intersections. NCHRP Report 233, TRB.

Webster, F.V., 1958. Traffic signal settings. road research technical paper, No.39, Great Britain Road Research Laboratory, London.