Available online at www.sciencedirect.com

ScienceDirect

RESEARCH PAPER

Timing Oversaturated Signals: What Can We Learn from Classic and State-of-the-art Signal Control Models

CHEN Shuaiyu[1], XU Hao[2], LIU Hongchao[2],*

1 Walter P Moore and Associates, Inc., 1301 McKinney Suite 1000, Houston, TX, USA
2 Dept. of Civil and Environment Engineering, Box 41023, Texas Tech University, Lubbock, TX, USA

Abstract: Many signalized intersections in the United States are operating in oversaturated conditions during the morning and afternoon peak hours. Unlike under-saturated intersections, for which many well-established theories and signal timing policies have been developed, consensus has not been reached with regard to the control philosophy and timing strategy of oversaturated signals. For an isolated intersection, a common belief is that timely and efficient allocation of green time among the intersection approaches is essential to optimize traffic flows. A couple of the studies in the past have resulted in reasonable control policies that could possibly be implemented. However, the advantages and limitations of these strategies remain unrealized due to a lack of effective evaluation. This article conducts an in-depth investigation to the characteristics and applicability of the control policies through thorough examination of two typical groups of classical models. The performance of the classic models is then compared with the latest versions of TRANSYT-7F and Synchro, upon which guidelines are made for effectively implementing signal control polices for oversaturated intersections.

Key Words: urban traffic; traffic signals; optimization models; oversaturated intersections

1 Introduction

More and more urban signalized intersections in the United States are operating in saturated or oversaturated conditions during the morning and afternoon peak hours. Unlike undersaturated intersections, for which many well-established theories and signal timing policies have been developed, consensus has not been reached with regard to the control philosophy and timing strategy of oversaturated signals. The majority of the traffic signals in the United States are timed by a handful of signal timing software packages (e.g., Synchro, TRANSYT-7F, and PASSER). How to improve their performance on timing oversaturated intersections is of significant importance to traffic engineers given the frequency of oversaturated conditions.

For an isolated oversaturated intersection, timely and efficient allocation of green time among the intersection approaches is essential to optimize traffic flows. Mathematical models on signal timing variables for oversaturated intersections have been proposed by a number of researchers. A couple of the studies were focused on the green time

allocation or phase switching policy and have resulted in reasonable control policies that could possibly be implemented. However, the implementation has not been seen in the field due in part to the fact that most of the models have never been evaluated.

Gazis and Potts[1] first revealed that the total intersection delay would be minimized if the queues on all (critical) approaches could be dissolved simultaneously. Gazis[2] further verified the concept by using an example of an oversaturated intersection of two one-way streets. Based on this principle, Michalopoulos and Stephnopolos proposed their so-called "bang-bang" two stage signal control model[3] in which a constraint on queue length was considered. Recent studies along this route include research of Chang and Lin[4,5], who proposed a discrete dynamic model and performance index approach to optimize signal parameters during the entire period of oversaturation. In the research of Yin, focuses were placed on investigating the effect of fluctuation on optimal timing of fixed-time signals[6].

Dans and Gazis[7] further stated that giving priority to the

DOI: 10.1016/S1570-6672(13)60093-8

oversaturated approach with higher saturation flow rate would be beneficial to delay optimization. In the same literature, a linear programming (LP) model was developed based on the so-called store-and-forward phase switching principle. Recently, researchers also started investigating control policies for a system of oversaturated intersections. For example, Lieberman et al.[8] conducted a mixed-integer linear program approach to explore feasible control policies of oversaturated arterials. Lo and Chow[9] developed a dynamic intersection signal control optimization method based on the cell-transmission model. Recent works along this line also includes work of Wong et al.[10]

The focus of another approach, which is different from the control principle/policy oriented study, is to develop control algorithms that rely on advanced computation, such as genetic algorithms (GA) for optimal signal timing and queue management of oversaturated intersections[11−13]. Li and Prevedouros presented a hybrid optimization and rule-based oversaturated control algorithm for adaptive signals[14]. The GA based approach seeks global optimum through the adaptive heuristic searching algorithm and does not necessarily rely on a clearly defined phase switching policy. A successful application of the GA based control model is TRANSYT-7F. In the latest version, Hale introduced their methodology for timing oversaturated signals[15].

Signal timing software packages have been widely used by traffic engineers and there are many advantages of using these readily available signal timing tools. However, the proprietary nature of the control algorithms embedded in those commercialized tools makes it difficult for in-depth investigation. On the other hand, the findings from theoretical models have not been used in practice to their full potential. This gap needs to be bridged for theory to remain relevant for practitioners. This in turn requires a thorough evaluation of the theoretical models and detailed comparison analyses with regard to the performance of the theoretical approaches and the prevalent signal timing tools.

This paper examines two groups of classic models and compares their performance with two prevailing signal timing tools, TRANSYT-7F, V 10.3 and Synchro 7. The first group includes the graphical model, Pontryagin's control model, and the discrete minimal delay model which belong to the bang-bang type of control and follow Gazis' simultaneous queue discharging principle; The second group includes the LP model which follows the principle of giving priority to oversaturated intersection approaches with maximum saturation flow rates; TRANSYT-7F was selected as a representative of the models whose green time allocation strategy is based on advanced formulation rather than a clearly defined phase switching policy; Synchro was selected to represent the models that assign green time according to the Highway Capacity Manual's method[16]. The findings of this

research may help researchers and practitioners in understanding the advantages as well as the applicability and limitations of these theoretical models. Understanding of the control principles, in return, may play a positive role in improving the existing signal timing tools.

2 Background

2.1 Oversaturated Conditions

A signalized intersection becomes oversaturated when traffic demand exceeds capacity, causing queues of vehicles to develop in one or all of the approaches of the intersection. The underlying assumption of the definition is that there is no effective green time loss at all intersection approaches during the oversaturated period. If one or more approaches have effective green time loss, the intersection is said to be in pseudo-oversaturated condition because by adjusting the signal plan the intersection may be converted back to under-saturated condition.

The effective green time loss is sometimes inevitable when traffic demand is reaching a certain threshold that a fluctuation has a reasonable chance of creating residual queues from one cycle to the next. Daganzo pointed out that such unpredictable stochastic fluctuations have little effect on the average delay if traffic is well under or well over saturation[17]. Throughout this paper, random delays caused by traffic fluctuation are not considered, only uniform delays are considered

Gazis and Potts[1] indicated that the sufficient condition for oversaturation at an intersection with a fixed time signal plan and constant cycle length C and lost time L per cycle is:

$$\sum_i \frac{v_i}{s_i} > 1 - \frac{L}{C} \qquad (1)$$

where v_i and s_i are the mean arrival rates and the saturation flow rates of the critical approach i of the intersection.

However, the current traffic condition is not only a result of current traffic arrivals and departures but also related to previous traffic conditions. At an oversaturated intersection, a certain amount of the effective green time must first serve the vehicles left during the previous cycle. In such a case, the steady-state condition breaks down and the signal timing task becomes more intricate than that for under-saturated conditions. Therefore the authors define the sufficient and necessary condition for oversaturation as:

$$\sum_i \frac{v_i + Q_i/C}{s_i} > 1 - \frac{L}{C} \qquad (2)$$

where Q_i is the number of vehicles in the residual queues on the critical approach i of the intersection.

2.2 Control Objectives and Constraints

The most often used control objectives for signal timing design and evaluation are minimum total intersection delay and/or maximum throughput of the intersection. The objective of minimum intersection delay or average vehicle delay has been arguably the sole criterion for signal timing of
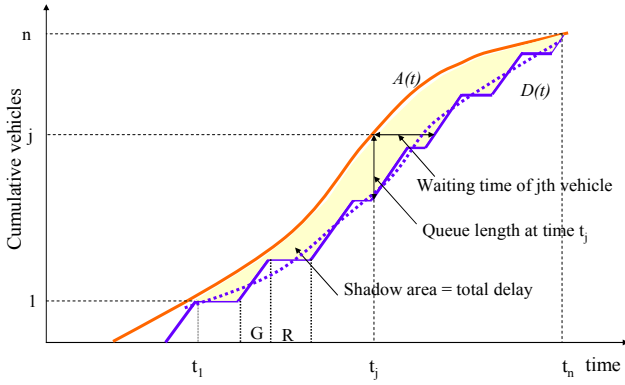
Fig. 1   Cumulative curves at an oversaturated approach

under-saturated intersections as used by Webster[18]. If stochastic delay can be neglected, the goal becomes maximizing total throughput from the intersection subject to the following constraints:

(1) Queue length; (2) Maximum individual delay (Maximum individual waiting time); (3) Minimum green time.

For a theoretical model to be of practical value, these constraints must be considered. For example, the objective of minimizing intersection delay can be easily achieved (in the model) if there is no constraint on the queue length. However, it may cause an unfavorable situation in which queue may spill back and block the upstream intersection. The restrictions on the maximum individual delay and the minimum green time are equally important from the practical point of view.

2.3   Cumulative Curve

A common objective for both fixed-time and real-time signal timing design is to develop traffic signal plans that minimize total vehicle delay, which can be achieved through the use of cumulative curves to represent (deterministic or uniform) traffic arrivals and departures. As shown in Fig. 1, many theoretical models construct cumulative arrival and departure curves to illustrate the queue formation and dissipation process during the congested period.

If we label the first vehicle when the queue starts as vehicle number 1, and the last vehicle to arrive when the queue disappears as number n, the shadowed area enclosed between $A(t)$ and $D(t)$ during the queuing time interval (t_1, t_n) represents the total vehicle delay d during the period of congestion.

$$d = \int_{t_1}^{t_n} [A(t) - D(t)]dt = \int_{t_1}^{t_n} Q(t)dt \qquad (3)$$

where $A(t)$ represents the cumulative number of vehicle arrivals on the approach over time; $D(t)$ represents the cumulative number of vehicle departures at the intersection over time; $Q(t)$ represents the number of vehicles stored in queues. The horizontal distance between $A(t)$ and $D(t)$ represents the individual vehicle delay before it discharges from the intersection. This time-dependent definition is helpful in developing online and adaptive signal timing plans, in which accurate projection of real-time arrival is the key[19].

If there are multiple lanes on one approach, in order to simplify the computation while maintaining the first-in-first-out (FIFO) criteria, one assumes that the queues are evenly distributed and no lane-changing behavior exists within the lanes. Delay of the approach is thus expressed by Eq. (4).

$$d_A = a d_L \qquad (4)$$

where d_A is delay of the approach; d_L is delay of single lane; a is the number of lanes on the approach.

Generally, departure curve $D(t)$ is transferred to the smooth approximation, the dotted line shown in Fig. 1. Hence, Eq. (5) shows the relationship between the queue length and the control variables.

$$\frac{dQ}{dt} = \frac{dA}{dt} - \frac{dD}{dt} = v - \frac{sg}{C} = v - u \qquad (5)$$

where v is the arrival flow rate; u is the departure flow rate; s is the saturation flow rate; g is effective green time; C is the cycle length.

2.4   Control Algorithms of the Selected Theoretical Models

2.4.2   The Graphical Approach

Gazis and Potts first revealed that the intersection delay could be minimized if the queues on all (critical) intersection approaches are cleared simultaneously[1]. Based on this principle, the control strategy was developed by a graphical approach. $g_{i\min}$ and $g_{i\max}$ are the minimum and maximum effective green time of the approach i respectively; and the associated minimum and maximum departure rates, i.e., $u_{i\min}$ and $u_{i\max}$, of approach i is:

$$u_{i\max,\min} = \frac{s_i g_{i\max,\min}}{C} \qquad (6)$$

If $S_1 > S_2$, the conceptual two-stage operation process (bang-bang control) is illustrated by Fig. 2.
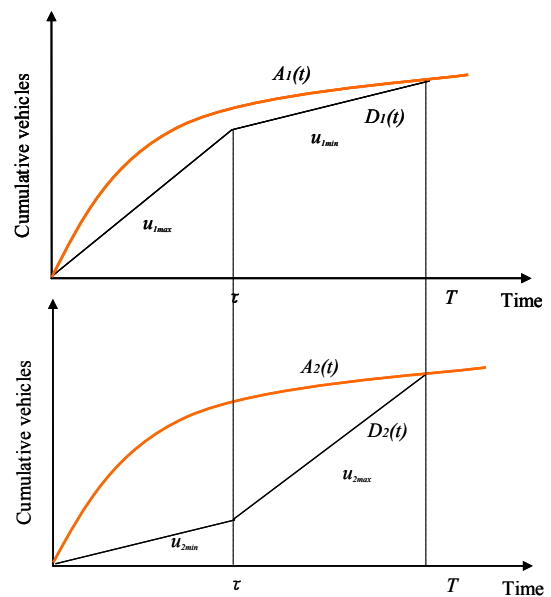

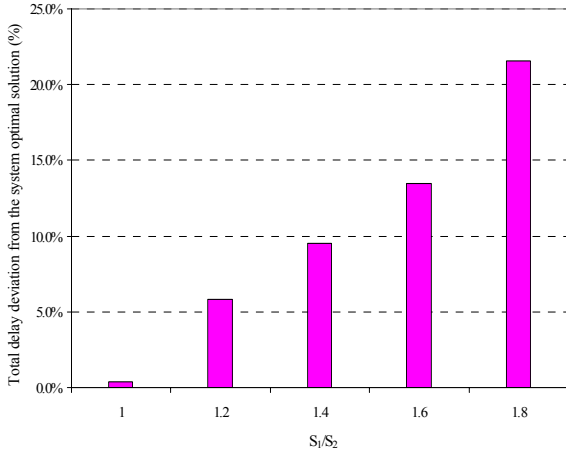Fig. 2   Optimal control of graphical model

Fig. 3 Total delay deviation under different saturation flow rates conditions

In the first stage, the departure rate of approach 1 is set to its maximum, u_{1max}, the departure rate of approach 2 is set to its minimum, u_{2min}; at the switch-over point τ, the second stage begins with the departure rate of approach 1 set to u_{1min}, and the departure rate of approach 2 set to u_{2max}. T is the termination time of the oversaturated period. Graphical model tends to search for the optimal switch-over point (τ) for minimal intersection delay. The solution is obtained by trial drawing and inspection.

2.4.2 Pontryagin's Control Model

Gazis[3] further developed his solution for oversaturated intersections by using Pontryagin's maximum principle. Michalopoulos and Stephnopolos placed a constraint on the queue length into Gazis model and established a discrete algorithm based on the numerical method. For an intersection with two one-way approaches, the objective is to minimize the total queue length, which is:

$$\min \int_0^T (Q_1(t) + Q_2(t)) \mathrm{d}t \tag{7}$$

where $Q_1(t)$ and $Q_2(t)$ is the queue length of the first and second approach, respectively and

$$Q_0(t) = \int_0^T (Q_1(t) + Q_2(t)) \mathrm{d}t$$

Steering functions f_0, f_1, f_2 were then defined for controlling the changes in the state variables:

$$f_0 = \frac{dQ_0}{dt} = Q_1 + Q_2$$

$$f_1 = \frac{dQ_1}{dt} = v_1 - u \tag{8}$$

$$f_2 = \frac{dQ_2}{dt} = v_2 - s_2(1 - \frac{L}{C}) + \frac{s_2}{s_1}u$$

where u (the departure rate of the first approach) is a function of time and called control variable, v_1, v_2 are the arrival flow rate for approach 1 and approach 2, respectively. The control domain is:

$$u_{min} \leq u \leq u_{max} \tag{9}$$

and the boundary conditions are:

$$Q_1(0) = Q_1(T) = 0$$
$$Q_2(0) = Q_2(T) = 0 \tag{10}$$
$$Q_0(0) = 0$$

The solution follows the classical Pontryagin's maximum principle. To implement Pontryagin's method one defines a Hamiltonian function:

$$H = \varphi_0 f_0 + \varphi_1 f_1 + \varphi_2 f_2$$

$$= \varphi_0(Q_1 + Q_2) + \varphi_1(v_1 - u) + \varphi_2(v_2 - s_2(1 - \frac{L}{C}) + \frac{s_2}{s_1}u) \tag{11}$$

where φ_1, φ_2, φ_0 are the functions of time. They are defined as follows:

$$\frac{d\varphi_1}{dt} = -\frac{\partial H}{\partial Q_1} = -\varphi_0$$

$$\frac{d\varphi_2}{dt} = -\frac{\partial H}{\partial Q_2} = -\varphi_0 \tag{12}$$

$$\frac{d\varphi_0}{dt} = -\frac{\partial H}{\partial Q_0} = 0$$

Therefore,

$$\varphi_0 = -1$$
$$\varphi_1 = t + K_1 \tag{13}$$
$$\varphi_2 = t + K_2$$

The optimum value of the control variable u at time t is the one that maximizes H.

$$\frac{\partial H}{\partial u} = (\frac{s_2}{s_1} - 1)t + (K_2\frac{s_2}{s_1} - K_1) \tag{14}$$

Thus, the control algorithm reads:

$$\frac{\partial H}{\partial u} > 0 \quad u = u_{max}$$

$$\text{if } \frac{\partial H}{\partial u} < 0 \quad u = u_{min} \tag{15}$$

$$\frac{\partial H}{\partial u} = 0 \quad \text{switch-over point}$$

This model has provided the theoretical solution of Gazis and Potts' graphical model and has verified the bang-bang control[1].

Michalopoulos and Stephnopolos further placed a queue length constraint into this model and drew the following conclusions: the optimal control u is affected by the queue length constraint during finite time interval[3]. When queue length reaches its constraint, the departure rate must be equal to the arrival rate. If queues on both approaches reach the constraints and the cycle length is constant, the problem may have no feasible solution.

2.4.3 Discrete Minimal Delay Model

Chang and Lin developed a discrete minimal delay model to solve the oversaturated problem[4]. Based on the delay model, a performance index model is also developed by introducing a penalty for the number of stops. In this paper, only their minimal delay model is considered. The objective function is to minimize total delay. The function is in quadratic form as below:

$$\min J = \frac{1}{2}(D(N))^2 + \frac{1}{2}\sum_{k=2}^N (D(k))^2 \tag{16}$$

where N is the terminative state of the oversaturated period.

Based on the optimal control theory, the problem is to minimize the Hamiltonian function as follows:

$$H = \frac{1}{2}(D(k))^2 + \lambda(k+1)[D(k) + Bu(k) + W(k)] \quad (17)$$

where $\lambda(k+1)$ is Lagrange multiplier in state $k+1$, B is the control gain; $u(k)$ is the control variable; $W(k)$ is exogenous variable.

For the isolated intersection with two one-way approaches:

$$u(k) = g_2^2(k)$$
$$B = \frac{1}{2}(a_1 s_1 - a_2 s_2) \quad (18)$$

where $g_2(k)$ is the effective green time of approach 2 in the k th cycle; a_1 and a_2 are the number of lanes on approach 1 and approach 2, respectively; s_1 and s_2 are the saturation flow rate for approach 1 and approach 2, respectively.

$$\frac{\partial H}{\partial u} = \lambda(k+1)B \quad (19)$$

To minimize H, the control process is:

$$\frac{\partial H}{\partial u} > 0 \quad g_2 = g_{2min}$$
$$\frac{\partial H}{\partial u} < 0 \quad g_2 = g_{2max} \quad (20)$$

Clearly, it is a bang-bang type control with a two-stage operation. The solution of this approach is consistent with the Pontryagin's control model.

2.4.4 Linear Programming (LP) Model

Dans and Gazis proposed their LP model for a system of oversaturated intersections, in which the problem of an oversaturated intersection was further addressed[7]. The model was developed based on the concept that priority is given to the intersection approaches with higher saturation flow rates when green time is allocated among signal phases. The key for this approach is to keep the subject intersection approach saturated before switching to next green phase. The model reads:

$$\min \int_0^T (Q_1 + Q_2) dt \quad (21)$$

$$\frac{dQ_1}{dt} = \begin{cases} v_1 - u & \text{if } Q_1 > 0 \\ \max(0, v_1 - u) & \text{if } Q_1 = 0 \end{cases}$$

$$\frac{dQ_2}{dt} = \begin{cases} v_2 - s_2(1 - \frac{L}{C}) + \frac{s_2}{s_1}u & \text{if } Q_2 > 0 \\ \max(0, v_2 - s_2(1 - \frac{L}{C}) + \frac{s_2}{s_1}u) & \text{if } Q_2 = 0 \end{cases} \quad (22)$$

$$u = \frac{s_1 g_1}{C} \quad (23)$$

$$g_{1min} \leq g_1 \leq g_{1max} \quad (24)$$

$$\frac{v_1 - \frac{dQ_1}{dt}}{s_1} + \frac{v_2 - \frac{dQ_2}{dt}}{s_2} \leq 1 - \frac{L}{C} \quad (25)$$

where T is the duration of the oversaturated period

$$[Q_1(t) = Q_2(t) = 0 \text{ for } t \geq T]$$

Consider N time intervals that $N \cdot \Delta t \geq T$ and define:

$$Q_i(n) \approx Q_i(n\Delta t), \quad i = 1, 2 \quad (26)$$

The objective function Eq. (21) becomes to:

$$\min \frac{1}{2} \sum_{n=0}^{N-1} \sum_{i=1}^{2} (Q_i(n) + Q_i(n+1))\Delta t \quad (27)$$

If one approximates the differential Eqs. (22) and (25) by difference equations, the problem becomes a standard LP problem.

As summarized in Table 1, the four classic models can be roughly classified into two groups by the patterns of queue dissipation. The first group includes the graphical model, Pontryagin's control model, and the discrete minimal delay model which follow Gazis' simultaneous queue discharging principle. By applying different optimal control methods, the models in this group lead to similar bang-bang type signal control with a two-stage operation. The second group includes the LP model which follows the principle of giving priority to oversaturated intersection approaches with maximum saturation flow rates. The effectiveness of these two groups of models are presented in the following section.

3 Comparison analyses of theoretical models

3.1 Minimum effective green time

The minimum green time is usually governed either by the pedestrian crossing time or the time to allow one or two queued vehicles to pass through the intersection for actuated signals. From the mathematical point of view, the minimum effective green time is a boundary condition. As mentioned earlier, the regulation of minimal effective green time is treated as a constraint during the control process. A simplified case is now presented to show the relationship between the regulation and the control algorithm.

Note that all of the models are concerned with the phase switching policy issue, using a simplified case with two one-way approaches and two phases would be sufficient. The following example is similar to that used in Michalopoulos and Stephnopolos[20]. Consider an intersection with two one-way approaches with a two-phase signal control and the cycle length is 150 s (C=150).

Table 1   Algorithms of theoretical oversaturated control models

	Graphical model	Pontryagin's control model	Discrete minimal delay model	LP model
Developer	Gazis and Potts (1963)	Gazis (1964), Michalopoulos and Stephnopolos (1977)	Chang and Lin (2000)	Dans and Gazis (1976)
Model type	Continuous	Continuous/Discrete	Discrete	Continuous
Objective	Minimal total delay	Minimal total delay	Minimal total delay	Minimal total delay
Constraint	None	Queue length	None	Queue length and maximum waiting time
Optimization method	Graphical method	Pontryagin's maximum principle	Linear quadratic method	Linear programming
Queue dissipation pattern	Simultaneously	Simultaneously	Simultaneously	Not simultaneously

Table 2　Five-minute cumulative arrival volumes

Time (s)	Approach 1 (veh/lane)	Approach 2 (veh/lane)
0	0	0
300	121	86
600	205	147
900	268	192
1,200	318	227
1,500	359	257
1,800	396	283
2,100	430	307
2,400	462	330
2,700	492	352
3,000	523	373
3,300	552	394
3,600	582	415
3,900	611	436
4,200	640	457

Table 3　Queue length and control strategy by bang-bang control strategy

Time (s)	Cycle No.	Approach 1		Approach 2	
		Green time 1	Queue length 1	Green time 2	Queue length 2
0	0	0.0	0	0.0	0
150	1	97.5	23	52.5	28
300	2	97.5	45	52.5	57
450	3	97.5	49	52.5	73
600	4	97.5	53	52.5	89
750	5	97.5	47	52.5	97
900	6	97.5	41	52.5	105
1050	7	97.5	28	52.5	107
1200	8	60.0	29	90.0	100
1350	9	60.0	26	90.0	90
1500	10	60.0	24	90.0	80
1650	11	60.0	19	90.0	68
1800	12	60.0	14	90.0	56
1950	13	60.0	8	90.0	43
2100	14	60.0	1	90.0	30
2250	15	60.0	0	90.0	16
2400	16	60.0	0	90.0	3
Total delay		60,819 (veh-s)		155,919 (veh-s)	

Approach 1 has one lane and s_1=1400 pcu/h, g_{1max}=0.65C (97.5 s), g_{1min}=0.4C (60 s); Approach 2 has one lane and s_2=1000 pcu/h, g_{2max}=0.6C (90 s), g_{2min}=0.35C (52.5 s). Table 2 gives the five minute's cumulative arrival volumes.

According to Michalopoulos and Stephnopolos and Chang and Lin[3,4], the optimal signal control scheme of the bang-bang control strategy is summarized in Table 3. The queues on both approaches are dissolved at roughly the same time (approach 1 at 2,100 s, approach 2 at 2,400 s). The switch-over point is at 1,050 s and the termination time is at 2400 s. The total intersection delay summed from approaches

1 and 2 is 216,738 veh-s, which is equivalent to 124.78 s/veh in terms of average vehicle delay. The signal scheme for the same problem calculated by the LP model is shown in Table 4.

The queue on approach 1 dissipates at 1,200 s and the rest of the green time is then allocated to approach 2. The queue on approach 2 dissipates at 2,550 s. Total intersection delay summed from approaches 1 and 2 is 224,789 veh-s, which is equivalent to 129.41 s/veh in average delay.

By comparing the results provided by these two types of models, it is revealed that the total delay calculated by the bang-bang control strategy is smaller than that of the LP model (216,738<224,789). In the LP model, when the queue on the approach with maximum saturation flow rate is dissolved, the signal needs to be maintained at the saturation state. However, if the arrival rate is smaller than the minimal service rate on that approach, there is effective green time loss. In this example, the queue on approach 1 is dissolved at 1200 s and queue on approach 2 is dissolved at 2,550 s. There is significant effective green time loss on approach 1 during [1,200 s, 2,550 s], which indicates that the intersection is in a pseudo-oversaturated condition. The constraint of minimal green time is shown in Eq. (28):

$$g_{1min} = u_{min}\frac{C}{s_1} \le \min(v_1(t)\frac{C}{s_1}) \quad (T_1 < t < T_2)$$
$$u_{min} \le \min(v_1(t)) \quad (T_1 < t < T_2)$$

(28)

where T_1 and T_2 are the queue duration time of approach 1 and 2, respectively.

Table 4　Queue length and control strategy of the LP model

Time (s)	Cycle No.	Approach 1		Approach 2	
		Green time 1	Queue length 1	Green time 2	Queue length 2
0	0	0.0	0	0.0	0
150	1	97.5	23	52.5	28
300	2	97.5	45	52.5	57
450	3	97.5	49	52.5	73
600	4	97.5	53	52.5	89
750	5	97.5	47	52.5	97
900	6	97.5	41	52.5	105
1050	7	97.5	28	52.5	107
1200	8	97.5	15	52.5	110
1350	9	90.5	0	59.5	109
1500	10	60.0	0	90.0	99
1650	11	60.0	0	90.0	87
1800	12	60.0	0	90.0	75
1950	13	60.0	0	90.0	62
2100	14	60.0	0	90.0	49
2250	15	60.0	0	90.0	35
2400	16	60.0	0	90.0	22
2550	17	60.0	0	90.0	7
Total delay		43900 (veh-s)		180,889 (veh-s)	

Table 5　Queue length and control strategy of system optimal solution

Time (s)	Cycle No.	Approach 1		Approach 2	
		Green time 1	Queue length 1	Green time 2	Queue length 2
0	0	0.0	0	0.0	0
150	1	150.0	2	0.0	43
300	2	150.0	4	0.0	86
450	3	119.1	0	30.9	108
600	4	108.0	0	42.0	127
750	5	81.0	0	69.0	130
900	6	81.0	0	69.0	133
1050	7	64.3	0	85.7	127
1200	8	64.3	0	85.7	121
1350	9	52.7	0	97.3	109
1500	10	52.7	0	97.3	97
1650	11	47.6	0	102.4	81
1800	12	47.6	0	102.4	66
1950	13	43.7	0	106.3	48
2100	14	43.7	0	106.3	31
2250	15	41.1	0	108.9	12
Total delay		975 (veh-s)		196,955 (veh-s)	

It is easy to verify that in bang-bang control, the control variable u_{min} must follow:

$$g_{1min} = u_{min}\frac{C}{s_1} > \min(v_1(t)\frac{C}{s_1})\,(0 < t < T)$$
$$u_{min} > \min(v_1(t))\quad(0 < t < T) \tag{29}$$

where T is the equalized queue duration time ($T=T_1=T_2$). Formula (28) and (29) indicate that the constraint of minimal effective green time has a significant impact on the performance of both groups, especially for the LP model because it is difficult to maintain the state of saturation flow in reality.

3.2　System optimal solution

If the constraint on the minimum green time is removed, the LP model provides a so-called system optimal solution in which the total intersection delay is further reduced. The solution of the same example is shown in Table 5. It is clear that the system optimal solution may not be practical since the drivers on approach 2 are punished too much. However, it provides an upper boundary threshold and can be used as a reference to evaluate other solutions.

3.3　Evaluation of the queue dissipation patterns

The major difference between the two types of models is the queue dissipation pattern. Gazis proposed the principle that queues should be dissolved simultaneously for minimum intersection delay[2]. However, the assumption Gazis made is that: "throughput is the same in both directions so that the total throughput of the intersection remains constant." The departure curves of two approaches D_1 and D_2 will thus be:

$$D_1 = Ft$$
$$D_2 = (K - F)t \tag{30}$$

where K is a constant and F is the average service rate on approach 1, and:

$$d_1 = \int_0^{T_1}[Q_1(t) - Ft]\mathrm{d}t$$
$$d_2 = \int_0^{T_2}[Q_2(t) - (K - F)t]\mathrm{d}t \tag{31}$$

The total departure of each approach at the end of congested period is:

$$Q_1(T_1) = FT_1$$
$$Q_2(T_2) = (K - F)T_2 \tag{32}$$

The total delay reaches minimum when:

$$\frac{d(Delay)}{dF} = \frac{d}{dF}(\int_0^{T_1}[Q_1(t) - Ft]\mathrm{d}t + \int_0^{T_2}[Q_2(t) - (K - F)t]\mathrm{d}t)$$
$$= \int_0^{T_2} t\mathrm{d}t - \int_0^{T_1} t\mathrm{d}t = 0 \tag{33}$$

Therefore, $T_2=T_1$ which indicates the queues will dissipate at the same time.

The fundamental assumption for the above result is that: "total throughput remains constant." If the total throughput of one cycle is defined as P, then:

$$P = s_1 g_1 + s_2(C - g_1 - L) \tag{34}$$

where L is the total lost time of the cycle and C is the cycle length. If the total throughput remains constant, then:

$$\frac{\mathrm{d}P}{\mathrm{d}g_1} = s_1 - s_2 = 0 \tag{35}$$

Equation (35) shows that the assumption Gazis made is valid only when $S_1=S_2$. In other words, dissolving queues simultaneously for delay minimization is valid only when the saturation flow rates of the (critical) approaches are equal or close. Therefore, the strategy should be only considered for the cases where the intersection approaches have same or similar saturation flow rates (e.g., intersections at two major streets). If the saturate flow rates of intersection approaches are significantly different with each other (e.g. where a major street intersects with a minor street), the strategy of bang-bang type control cannot provide the optimal solution. This finding is of particular significance to traffic engineers when applying the Gazis's principle to practice.

Applying the Gazis's principle to other types of intersections cannot lead to an optimal solution. For the previous example, we keep the saturation flow rate of approach 2 constant (s_2=1000 pcu/h) and calculate the total delay by the bang-bang control strategy and compare the result with the system optimal solution with different saturation flow rates on approach 1 (s_1=1000, 1200, 1400, 1600, 1800 pcu/h). By using the system optimal solution as a reference, the bang-bang control strategy becomes less effective with the increase of the difference in saturation flow rate.

The difference can also be visualized by inspecting Fig.3, which compares the bang-bang control with the system optimal solution by changing the saturation flow rates of the two intersection approaches. For instance, when S_1/S_2=1.8, compared to the system optimal solution, the bang-bang control strategy increases total delay by about 21.6%.

4   Evaluation of Synchro and TRANSYT-7F

Synchro and TRANSYT-7F are two widely used offline traffic signal optimization models. Both models have proven to be powerful tools when they are applied to undersaturated conditions. Recent development of the models also provides, to varying extent, strategies for timing oversaturated signals. Synchro and TRANSYT-7F are the models of choice because they represent the other two popular approaches in green time optimization. Synchro assigns green time among signal phases by using the Highway Capacity Manual's method, which was basically designed for undersaturated signals; TRANSYT-7F's green time allocation strategy is based on advanced formulation rather than a clearly defined phase switching policy. A comparison analysis of these two commercial tools and the classic models would be helpful for this study in assessing potential benefits of improved methods. As the focus of this paper is on isolated intersections, only their green time allocation/splits strategies are evaluated.

4.1   Synchro

Synchro is an offline signal optimization program which implements the Intersection Capacity Utilization (ICU 2003) and Highway Capacity Manual 2000 (HCM 2000) for intersection capacity analysis and signal timing design. From Synchro 6, congested link analysis and optimization was included in the software package[21]. A new total delay concept which includes the traditional control delay plus a new term of queue interaction blocking delay was embedded into Synchro 6. However, although the delay function was enhanced in terms of optimized cycle length and offset, there seems no improvement made on splits optimization.

In the latest version of Synchro 7, green time is allocated based on the flow ratio, which is the ratio of the flow rate and the saturation flow of the critical movement[21]. The split optimizer respects the minimum split setting for each phase whenever possible. When optimizing splits, it first attempts to provide enough green time to serve the 90th percentile lane group flow. If there is no enough cycle time to meet this objective, Synchro attempts to serve the 70th percentile traffic and then the 50th percentile traffic. Extra green time, if any, is given to the main street. In under-saturated conditions, this method tends to give lower v/c ratios (Eq. (36)) to the lane groups with low traffic demands. When traffic volume exceeds capacity, Synchro attempts to balance v/c ratios for each phase, while still respecting all minimums.

$$x = \frac{v}{s \times g / C} \tag{36}$$

It is clear that in under-saturated conditions, Synchro's policy for green time split is based on the adjusted Webster method which favors the main street phases. It is worthy of note that Synchro 5.0, 6.0, and 7.0 uses the same strategy for oversaturated conditions. Furthermore, because of its single

period optimization method, Synchro cannot model the accumulation of residual queues from one cycle to the next in oversaturated conditions.

4.2   TRANSYT-7F

TRANSYT-7F is another popular signal timing optimization tool, which applies genetic algorithm to optimize cycle length, phasing sequence, splits, and offsets. In its latest release, Version10.3[15], a multi-period optimization process was added. The purpose of multi-period optimization is to develop an effective signal-timing plan for handling oversaturated conditions and residual queues. The program can optimize the splits and offsets of one timing plan across all time periods. Genetic algorithm (GA) optimization is a theoretical improvement over the traditional hill-climb optimization technique that has been employed by T7F for many years. It is advantageous in searching the "global optimum" without being trapped in "local optimum" solutions.

4.3   Evaluation

The example used in section 4 is carried out to evaluate the performance of Synchro 7.0 and TRANSYT-7F Version 10.3, the results are shown in Figs. 4 and 5. Fig. 4 illustrates the performance of Synchro, TRANSYT-7F, and the theoretical models in terms of the total intersection delay which is presented by the area between the arrival and departure curves.
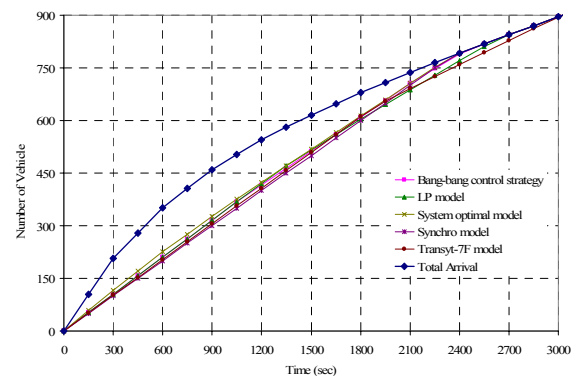

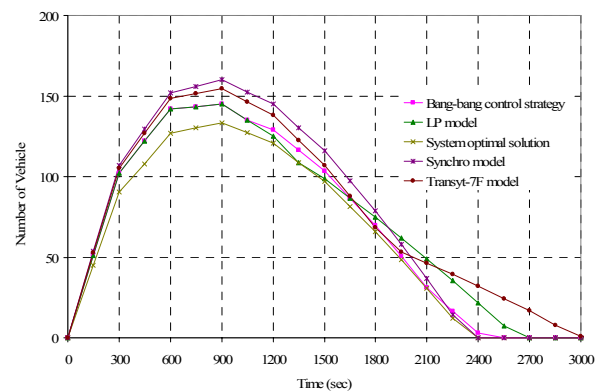Fig. 4   Comparison of delay by cumulative curves

Fig. 5   Comparison of queue length

For oversaturated intersections, the length of queue needs to be carefully managed because the queues blocking back to the upstream intersection may cause serious damage to the whole system. Fig. 5 compares the models' performance in terms of the queue length in number of vehicles.

It is not surprising that the system optimal model stands out from the others in both categories. But our earlier analysis has shown that this approach is not practical because of its extremely imbalanced green time allocation strategy. Here, it is used again as a reference. The bang-bang type of control performs fairly well, which is only next to the ideal solution from the system optimum model in both categories. The intersection delay from Synchro is slightly higher than that from the bang-bang strategy but the queue clearance time is almost the same. The LP model is next to the above three models due to its green time loss as described earlier. The TRANSYT-7F model, interestingly, performs least effectively in terms of intersection delay due to its relatively longer queue clearance time.

As for the queue length control, the three theoretical models outperform Synchro and TRANSYT-7F. As can be seen from Fig. 5, the maximum number of vehicles in queue resulted from Synchro and TRANSYST-7F is over 150, which is higher than that from the three theoretical models.

In summary, the findings from the evaluation can be summarized as follows:

(1) The models with clearly defined phase switching or green time allocation policies (the bang-bang control models, the LP model, and Synchro) perform better than the formulation based approach TRANSYT-7F in both intersection delay and queue clearance time. (2) The classic models with policies specifically designed for the oversaturated conditions (the bang-bang control model and the LP model) are advantageous in queue management compared to Synchro, which uses the same policy for under-saturated signal design, and the formulation based TRANSYT-7F. (3) The bang-bang type control models developed on the basis of the simultaneous queue dissipation philosophy are applicable only to the situation that the saturation flow rates of (critical) intersection approaches are equal or close. In addition, traffic demands during the whole congested period must be given in advance, which makes the bang-bang control strategy difficult to implement in practice, especially for real-time applications. (4) The LP model can lead to the system optimal if the constraint on minimum green time is relaxed and the condition of saturation is satisfied at all intersection approaches during the entire congested period. The solution can be taken as a reference to scale the performance of the other models.

5   Conclusions

The authors examined two groups of classic models and compared their performance with TRANSYT-7F and Synchro in signal timing of isolated and pretimed intersections. Focuses were placed on their green time allocation or phase switching policies under oversaturated traffic conditions. The bang-bang type of models follow the policy that queues are discharged simultaneously on all (critical) intersection approaches; the LP model is consistent with the strategy that priority is given to the oversaturated approaches with higher saturation flow rates; Synchro follows the principle that green time is allocated based on the flow ratio of critical intersection approaches. These three models all allocate green time according to a clearly defined control policy. TRANSYT-7F, on the other hand, features signal timing optimization through advanced computation which does not rely on a predefined policy.

The analyses conducted in this research revealed the necessity of embedding the classical models (with clearly defined phase switching policy) with the existing signal timing tools. When these integrated tools are applied, however, particular attention should be paid on the applicability of these models for best performance. For practical application of the results concluded from this research, the following guidance can be used for timing isolated oversaturated intersections.

(1) Intersections need to be classified based on their geometric layout (in terms of number of lanes at intersection approaches and associated saturation flow rates) and different type of intersections needs to be treated differently.

(2) For the intersections with asymmetric layouts (e.g. the intersections crossed by a major and a minor road), the policy of giving priority to the major street with higher saturation flow rate should apply and the LP model can be used. Particular attention should be paid to two points for best performance: a) the minimum green interval of minor streets needs to be carefully designed to avoid effective green time loss; and b) the growth of residual queue needs to be restrained according to the storage capacity of the link to avoid queue spilling back to the upstream intersection.

(3) For the intersections with symmetric layout and without significant difference in saturation flow among (major) intersection approaches (e.g. the intersections crossed by similar class of roads, minor & minor or major & major), the policy of simultaneous queue discharging and bang-bang type of models can be applied if the traffic demands during the entire congested period is available in advance. Otherwise, the policy of assigning green time according to flow ratio (i.e. Webster's method) would be sufficient and Synchro type of tools will work.

References

[1]   Gazis D C, Potts R B. The Oversaturated intersection. Proc., 2nd Int. Symp. on Theory of Road Traffic Flow, Paris, France, 1963, 221–237.

[2]   Gazis D C. Optimal control of a system of oversaturated

intersections. Operations Research, 1964, 12: 815–831.

[3]   Michalopoulos P G, Stephnopolos G. Oversaturated signal systems with queue length constraints-I single intersection. Transportation Research 11, 1977, 413–421.

[4]   Chang T H, Lin J T. Optimal Signal timing for an oversaturated intersection. Transportation Research, Part B, 2000, 34(6): 471–491.

[5]   Chang T H, Lin J T. Optimal Signal timing for an oversaturated intersection. Transportation Research, Part B, 2004, 38(8): 687–707.

[6]   Yin Y. Robust optimal traffic signal timing. Transportation Research Part B, 2008, 42: 911–924.

[7]   Dans G, Gazis D C. Optimal control of oversaturated store-and forward transportation networks, Transportation Science, 1976, 11(1): 1–19.

[8]   Lieberman E B, Chang J, Prassas E S. Formulation of real-time control policy for oversaturated arterials. Transportation Research Record 1727, Transportation Research Board, Washington, D.C., 2000, 77–98.

[9]   Lo H K, Chow A H. Control strategies for oversaturated traffic. J. of Transp. Eng., 2004, 130(4): 466–478.

[10] Won C K, Wong S C, Lo H K. A special queuing approach to optimize coordinated signal settings to obviate gridlock in adjacent work zones. Journal of Advanced Transportation, 2010, 44: 231–244.

[11] Park B, Messer C J, Urbanik II T. Traffic signal optimization program for oversaturated conditions: genetic algorithm approach. Transportation Research Record 1683, Transportation Research Board, Washington, D.C., 1999, 133–142.

[12] Park B, Messer C J, Urbanik II T. Enhanced genetic algorithm

for signal-timing optimization of oversaturated intersections. Transportation Research Record 1727, Transportation Research Board, Washington, D.C., 2000, 32–41.

[13] Abu-Lebdeh G, Benekohal R F. Genetic algorithm for traffic signal control and queue management of oversaturated two-way arterials. Transportation Research Record 1727, Transportation Research Board, Washington, D.C., 2000, 61–67.

[14] Li H, Prevedouros P D. Traffic Adaptive control for oversaturated isolated intersections: Model development and Simulation testing. J. of Transp. Eng., 2004, 130(5): 594–601.

[15] Hale D. Traffic Network Study Tool– TRANSYT-7F. McTrans, University of Florida, Gainesville, Florida, 2006.

[16] Transportation Research Board (TRB). Highway Capacity Manual 2000[M], Transportation Research Board, Washington, D.C, 2001.

[17] Daganzo C F. Fundamentals of Transportation and Traffic Operations. Pergamon, New York, 1997.

[18] Webster F V. Traffic signal settings. Road Research Technical Paper, No. 39, Road Research Laboratory, London, 1958.

[19] Fang F C, Elefteriadou L. Modeling and simulation of vehicle projection arrival-discharge process in adaptive traffic signal controls. Journal of Advanced Transportation, 44(1010) 176–192.

[20] Michalopoulos P G, Stephnopolos G. Optimal control of oversaturated intersections theoretical and practical considerations. Transportation Engineering & Control, 1978, 5: 216–221.

[21] Husch D, Albeck J. Trafficware SYNCHRO 7 User Guide, Trafficware, Albany, California, 2004.