# Development of an Optimization Methodology for Adaptive Traffic Signal Control at Diamond Interchanges

Fang Clara Fang[1] and Lily Elefteriadou[2]

**Abstract:** This research develops a methodology and a corresponding implementation algorithm to provide optimal signal control of diamond interchanges in response to real-time traffic fluctuations. The problem is formulated as to find a phase sequencing decision with a phase duration that makes a prespecified performance measure minimized over a finite horizon that rolls forward. The problem is solved by a forward dynamic programming (DP) method. The optimal signal switches over each 2.5 s interval are found for each horizon of 10 s. The optimization process is based on the advanced vehicle information obtained from loop detectors set back a certain distance from the stop line. Vehicle trajectories from detections till future arrivals and departures is modeled at the microscopic level to estimate the traffic flows at the stop-line for each horizon. The DP algorithm is coded in C++ language and dynamically linked to AIMSUN, a stochastic microsimulation package, for evaluation. The simulation results have exhibited that the DP algorithm is superior to PASSER III and TRANSYT-7F in handling demand fluctuations for medium to high flow scenarios when the field demand is increased from the one used in off-line optimization. The performance of the three algorithms is almost identical if the simulation demand is similar to off-line demand situation and does not vary much.

**DOI:** 10.1061/(ASCE)0733-947X(2006)132:8(629)

**CE Database subject headings:** Optimization; Methodology; Traffic signals; Interchanges; Traffic control; Simulation; Algorithms.



## Introduction

The signalization of two closely spaced intersections in diamond interchanges presents a major challenge in providing efficient traffic operations within the highway system. As shown in Fig. 1, the distance between the two intersections (D) varies from less than 122 m (400 ft) in densely developed urban areas to 244 m (800 ft) or more in suburban areas (Messer and Bonneson 1997). The close proximity of the two intersections creates a number of interactive effects that complicate the operation. This distance limits the storage available for queued vehicles. If the signal timing is not properly set, the queues from the downstream intersection would spill back and block the upstream approaches. Another phenomenon that adversely affects interchange operation is called demand starvation. It occurs when portions of the green at the downstream intersection are not used because conditions prevent vehicles at the upstream intersection from reaching the downstream stop line (Transportation Research Board 2000). Additionally, diamond interchanges usually have predominant left-turn movements that unfavorably affect progression and platoon cohesion along the surface street. To mitigate these unique operational problems at diamond interchanges, it is essential to provide optimal signal control of these two closed-space signalized intersections for efficiently accommodating all movements involved.

The most common diamond interchange timing plans are three phase or four phase. Each of them has advantages and disadvantages in minimizing delay and queuing, but none of them is optimal for every possible diamond interchange geometry and traffic pattern. From the literature review undertaken, the only existing optimization model for diamond interchanges is PASSER III, developed by Texas Transportation Institute (TTI) (Messer et al. 1977; Fambro et al. 1991). It optimizes the pretimed signal plan based on off-line demand and cannot adapt itself to fluctuating demand situations. Moreover, its optimal signal plan is chosen among a restricted number of alternatives and it is not a globally optimum solution. Therefore, there is a need to develop an adaptive and optimal signal control of diamond interchanges by considering real-time traffic flow conditions.

The main objectives of this research are to develop a methodology and the corresponding implementation algorithm to provide optimal signal control of diamond interchanges in response to real-time traffic fluctuations. The performance of the proposed optimal algorithm is evaluated by microsimulation. To achieve these objectives, the forward dynamic programming (DP) is applied as the optimization method to achieve global optimization, vehicle trajectories from detections till future arrivals and departures is developed at the microscopic level based on loop detectors set back a certain distance from the stop line, and a dynamic communication between the DP algorithm and the AIMSUN, a stochastic microsimulation package, is built for the evaluation. To enable the algorithm to implement practical scenarios, a majority rolling technique is also proposed in this study. The research focuses on using the three-phase diamond interchange

[1]Assistant Professor, Dept. of Civil and Environmental Engineering, Univ. of Hartford, West Hartford, CT 06117 (corresponding author). E-mail: fang@hartford.edu

[2]Associate Professor, Dept. of Civil and Coastal Engineering, Univ. of Florida, Gainesville, FL 32611. E-mail: elefter@ce.ufl.edu; formerly, Director, Transportation Research Center.

Note. Discussion open until January 1, 2007. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on May 3, 2005; approved on November 8, 2005. This paper is part of the ***Journal of Transportation Engineering***, Vol. 132, No. 8, August 1, 2006. ©ASCE, ISSN 0733-947X/2006/8-629–637/$25.00.

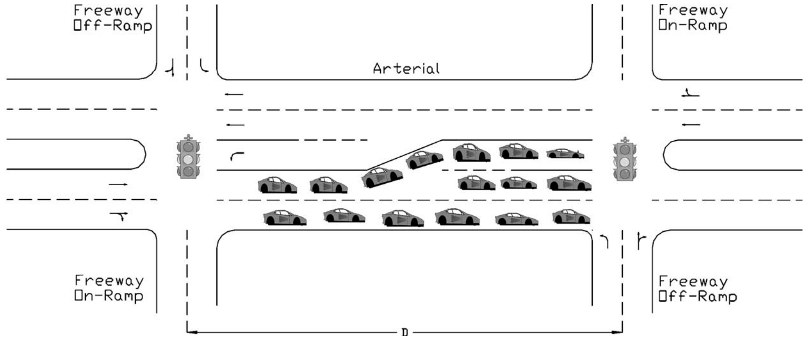


**Fig. 1.** Diamond interchange configuration



ring structure as an example, without loss of generality.

The following section gives an overview of past research on adaptive signal control systems. "Optimization Methodology" discusses the DP optimization methodology and vehicle trajectories model. The decision network, DP formulation, performance measures as well as fixed and dynamic weights are also discussed. "Algorithm Implementation" describes the algorithm implementation, which is followed by a short description of simulation procedure in "Simulation Evaluation of Developed Algorithm." "Comparisons" presents a comparison study on the performance of three different optimization methods, i.e., the proposed DP algorithm, PASSER III, and TRANSYT-7F. Conclusions and recommendations are provided in the last section.

## Literature Review

This section presents a brief literature review and is focused on adaptive signal control systems only. A more comprehensive and critical review on signal control and optimization of intersections and diamond interchanges is included in Fang (2004).

Adaptive control generates and implements the signal plan dynamically based upon real-time traffic conditions, which are measured through a traffic detection system. The algorithm for developing the signal plan differs widely between the various adaptive control systems. Compared to pretimed and actuated control, adaptive signal control has the potential to increase the operational efficiency of existing roadways, particularly for highly fluctuating demand, but undersaturated traffic conditions. The adaptive signal control concept was initiated by Miller (1963). He described an algorithm for adjusting signal timings in small time intervals of 1–2 s. A decision to be made is whether to extend the current green duration or terminate it immediately. The algorithm calculates the difference in vehicle-seconds of delay between the gain made during an extension and the loss in the cross street resulting from that extension. Since Miller's pioneering work, some research has been done in the United Kingdom, Australia, and the United States to develop adaptive systems. Systems such as SCOOT (Hunt et al. 1982), SCATS (Luk 1984; Charles 2001), OPAC (Gartner 1983; Gartner and Pooran 2002), and RHODES (Sen and Head 1997; Mirchandani and Head 2001; Mirchandani and Lucas 2001) are among the best known. Some of them were tested and implemented in many cities worldwide.

Among the aforementioned adaptive control systems, OPAC and RHODES are two major developments that are based on dynamic programming and operate using the rolling horizon concept. They have been developed primarily for individual intersections and are being extended to networks. For brevity, this section focuses only on the fundamental differences between those algorithms and the DP algorithm proposed in this paper. OPAC does not use a dynamic programming solution procedure, but rather than a restricted optimal sequential constrained search (OSCO). Consequently, the OPAC optimization methodology should be classified as a trial-and-error enumeration. Also, this methodology cannot be applied to optimize phase sequence. The DP algorithm is different from RHODES in two aspects: the dynamic programming formulation and the vehicle arrival prediction model. RHODES input includes the phase sequence, and during the optimization some of the phases may be skipped. However, in the DP methodology the phase sequence is an output of the optimization. Second, the RHODES formulation uses a longer horizon of about 60 s and applies a PREDICT model (Head 1995; Mirchandani and Head 2001). The DP algorithm uses a 10 s optimization horizon. Using a shorter horizon is likely to result in more accurate prediction of vehicle arrivals.

## Optimization Methodology

### *Overview*

Fig. 2 presents a NEMA phasing for diamond interchange movements. Using the NEMA phase numbering, as shown in Fig. 3, the adaptive signal control of diamond interchanges is formulated as a decision network problem in this study. DP is applied as the optimization method to find an optimal decision trajectory from the network since the DP is a highly effective and efficient mathematical technique for making a sequence of interrelated decisions. The optimal decision path consists of phase sequencing decision with phase duration that makes a prespecified performance measure minimized over a finite horizon that rolls forward. The optimization performance measure considers delay, queue length or queue storage ratio. Storage ratio is calculated as the queue length to the available space for queue storage. A horizon of 10 s is divided into an integer number of intervals, each having 2.5 s. The algorithm finds the optimal signal switches by minimizing the selected performance index over each horizon, and the optimization process proceeds one horizon after another. This DP problem is solved by forward value iterations in the algorithm. As the DP optimization operates on advanced vehicle information over each horizon, a vehicle trajectories model of future vehicular detections, arrivals, and departures is developed at the microscopic level based on loop detectors set back a certain distance



J. Transp. Eng., 2006, 132(8): 629-637

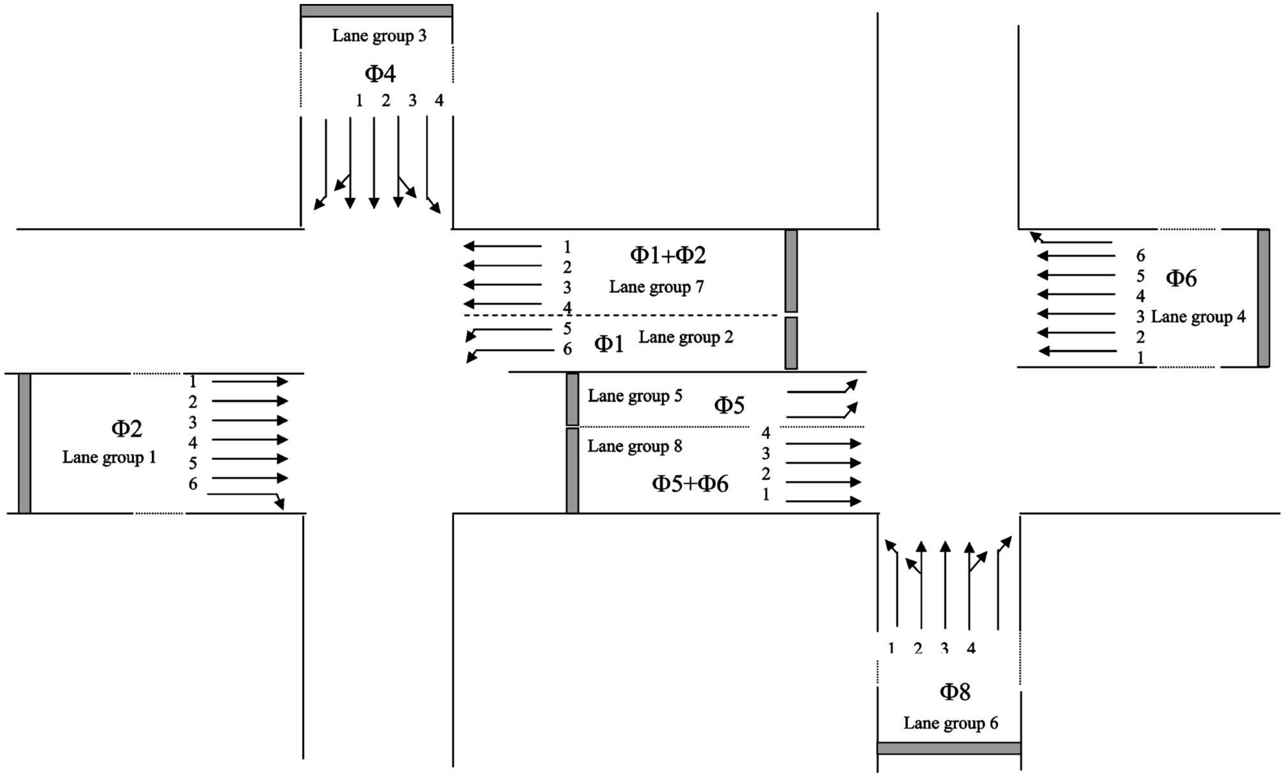


**Fig. 2.** NEMA phasing and detectors placement layout

from the stop line. This model accounts for the discrete-event dynamics of the signal switches, vehicle arrivals and discharges. With this model the traffic flows at the stop line can be precisely predicted for a horizon that the DP optimization operates on.

### *Decision Network and Forward DP Formulation*

The adaptive signal optimization problem for a diamond interchange is represented by a decision network associated with each optimization horizon (Fig. 3). A decision network is a network of interconnected nodes that are arranged in layers. A node with a label in it is the "state" in the DP formulation and also represents green signal phase status for certain movements. For example, Node 26 represents that the signal green phase for both EB and WB external links. The optimization horizon is defined as 10 s. The time is discretized at four constant stages, i.e., Stage 1, Stage 2, Stage 3, and Stage 4. Each stage stands for 2.5 s. At each discrete time $(t1, t2, t3\ldots)$ there are several possible signal Phases, i.e., Phases 26, 25, 15, 16, or 48. Thus a layer can have a total of five nodes representing possible signal phases at a time. Single input to a node is the decision coming from one of the previous possible phase, and multiple outputs departing from a node represent different signal choices at next stage. The unique

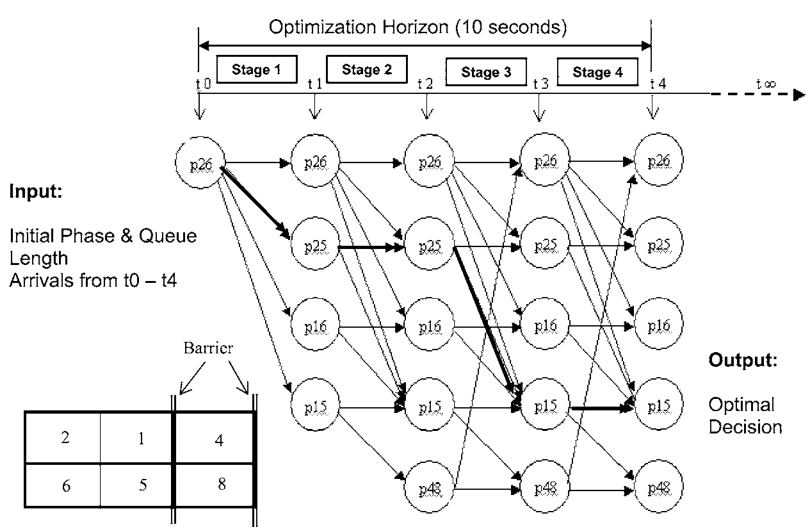


**Fig. 3.** Adaptive signal control in decision network solved by forward DP using three-phase ring structure as example



J. Transp. Eng., 2006, 132(8): 629-637

**Table 1.** Terminology Definition and Symbols for Forward DP Formulation

| Terminology | General DP definitions (Hastings 1973) | Definitions for forward DP adaptive signal optimization (Refer to Fig. 4) | Symbol |
|---|---|---|---|
| Stage | The original problem is divided into $N$ stages | An interval of 2 s, for example $n$=index of stage, also the number of stages from the Stage 0 to current stage | $n$ |
| State | Each stage has a number of states associated with it. The states are the various possible conditions in which the system might be at each particular stage of the problem | Signal phase state $\boldsymbol{i}$ for stage $n$ $i=1,2,\ldots,9$, for phases $P_{15}$, $P_{16}$,$P_{18}$, $P_{25}$, $P_{28}$, $P_{45}$, $P_{46}$, and $P_{48}$, respectively | $(n,i)$ |
| Decision variable | At each state there is a set of decisions from among which a choice must be made | Decision $k$ is a signal phase leading to the state $i$ at stage $n$. The decision departs from the state $j=k$ at stage $n-1$. $k$ is on decision space$K_i=\{1,2,\ldots,9\}$ $=\{P_{15}$, $P_{16}$, $P_{18}$, $P_{25}$, $P_{26}$, $P_{28}$, $P_{45}$, $P_{46}$, and $P_{48}\}$ | $k$ |
| State transition | A function that shows the state for the next stage changes based on the current state, stage, and optimal decision | Signal state transition $j=f(n,i,k)=k$ | J |
| Immediate return | Something which the system generates over one stage given the starting state and the decision | Queue length, delay, or user-defined performance measures over stage $n$, due to state $(n-1,j)$, queue length $q$ and decision $k$ (leading to state $i$ at stage $n$). The defined performance measure(s) is named as performance measure index (PMI). It is the sum of the PMI of eight lane groups in a diamond interchange. | PMI |
| Recurrence equation | Identifies the optimal policy at stage $n$, given that the optimal policy at stage $n-1$ is available | Eq. (1) | — |
| Optimal value of state | Best total value from stage 0 to stage $n$, given that the starting state at stage 0, and a sequence of optimal decision is made | Minimum PMI at a given state, from stage 0 up to stage $n$ under an optimal plan | $f^{*}(n,i)$ |

representation of the decision network make it possible to optimize both signal phase sequence and phase duration simultaneously.

The decision network represents all possible signal switch options. This research uses the basic three-phase diamond interchange ring structure (Fig. 3) as an example. A signal plan is a decision trajectory in the network that is composed of one node in each layer. The forward DP formulation procedure suggested by Hastings (1973) is adopted in this study to find the optimal decision path. Its terminology, definition and symbols are summarized in Table 1. It is expressed descriptively as: (1) The minimal performance measure index (PMI) value from stage 0 to stage $n$, at state $(n,i)$ is determined by finding the minimum value over all decisions $k$ for the sum of minimum (PMI) value from Stage 0 to stage $n-1$, at state $(n-1,j)$ and (PMI) value for state $(n-1,j)$ leading to state $(n,i)$ with decision $k_i$ and (2) mathematically as

$$f^{*}(n,i) = \underset{k\in K_i}{\mathrm{Min}}[\mathrm{PMI}(n,i,k,q) + f^{*}(n-1,j)] \qquad (1)$$

where $n$=stage number; $i$=current state at stage $n$ (or signal phase over the stage $n$); $q$=initial queue lengths at stage $n$. It is a vector of queue length of individual lane group as components for totally 8 lane groups; $k$=signal phase switch decision at stage $n-1$ that leads to state $i$ at stage $n$; $k_i$=set of possible decisions leading to state $i$; $j$=state at stage $n-1$ (or signal phase over the stage $n-1$); PMI $(n,i,k,q)$=return over stage $n$, due to decision $k$, state $(n-1,j)$ changing to state $(n,i)$, given initial queue lengths at stage $n-1$;$f^{*}(n-1,j)$=minimum PMI values from Stage 0 to state $n-1$; and $f^{*}(n,i)$=minimum PMI values from stage 0 to state $n$.

Accordingly, the basic DP terminology including the stage (2.5 s) and state (signal green phase) is defined in the context of the signal optimization problem, together with the symbols used to denote the relevant quantities. The equations that are needed for the solution of the optimal decision trajectory are also formulated (refer to Fang (2004) for details). They include:

- State transition equation;
- Immediate return equation;
- Optimal value function; and
- Terminal values.

In summary, the network nodes stand for phase states, and the arrowed lines for possible decisions and transitions of signal status. The DP optimization solution procedure involves the development of state transition function over a stage and the optimization of the objective function over an optimization horizon. The inputs to the optimization are traffic arrivals on individual lane groups, together with initial signal phase and initial queue length while outputs are signal phases at discrete times over the entire optimization period. Traffic arrivals are predicted from upstream detectors.

## Optimization Objective

The global optimal solution, i.e., optimal decision trajectory, is pursued with respect to a certain objective function that specifies operational performance of a diamond interchange. This research defines a PMI as this function. The performance measures considered include queue length, delay, and storage ratio. Using queue length as an example, Eq. (2) shows that PMI is the sum of



J. Transp. Eng., 2006, 132(8): 629-637

the average queue length on each of eight movements of an interchange and calculated over a horizon that consists of four stages

$$\mathrm{PMI} = \sum_{j=1}^{4} \left( \sum_{i=1}^{8} w[i]^{*}\mathrm{Qfinal}[i] \right)_j \tag{2}$$

where $w[i]$=weight associated with movement $i$; and Qfinal$[i]$=average queue length per lane on movement $i$ in the end of a DP interval $j$.

The PMI computed over each stage is also called an immediate return in the DP formulation. Given an initial queue length and signal phase in the beginning of a stage, the immediate return (e.g., queue length) at the end of a stage is dictated by the state transition function (i.e., a function about how the signal state is switched) that is based on arrival–discharge dynamics, and the decision to extend or switch from the initial phase. Therefore, a PMI objective function for optimal signal control is a function of signal phase, traffic condition, and state transition function. By minimizing a PMI value over entire optimization horizon of 10 s, the optimal decision on each interval can be located using the forward DP approach.

### Fixed Weights and Dynamic Weights

The weights $w[i]$ associated with movement $i$ in the PMI formulation [Eq. (1)] can be defined as fixed weights or dynamic weights. Fixed weights mean that their value associated with a lane group is predetermined and remains constant throughout the simulation, as illustrated in Eq. (2). The value will not change during the simulation. Dynamic weights dw$[i]$, as shown in Eq. (3), denote that the values of weights associated with each lane group vary in each optimization horizon during the simulation

$$\mathrm{PMI} = \sum_{j=1}^{4} \left( \sum_{i=1}^{8} \mathrm{dw}[i]^{*}\mathrm{Qfinal}[i] \right)_j \tag{3}$$

where dw$[i]$=dynamic weight associated with movement $i$; and Qfinal$[i]$=average queue length per lane on movement $i$ in the end of a DP interval $j$.

Dynamic weights are proposed in this research to be used in high demand (congested) and unpredicted conditions to quickly discharge queues for certain lane group(s). The values of dynamic weights are determined based on the storage ratio of a movement, as shown in Table 2. When the storage ratio of a movement is greater than certain values (e.g., 0.6), a higher dynamic weight value (e.g., 6) will be automatically assigned to this movement to reduce queue length. Table 2 also presents an example where ramps and arterials have a different specification of dynamic weight values. Ramps are assigned heavier weights than other lane groups under the same storage ratio. This takes into account that the current DP decision network (with three-phase ring structure) used does not provide enough ramp phase options.

**Table 2.** Dynamic Weight Values

| Movement type | Dynamic weight value |
|---|---|
| Arterial (EB and WB) | if SR$[i]>0.1$, then dw$[i]$=1.0 |
| | if SR$[i]>0.2$, then dw$[i]$=2.0 |
| | if SR$[i]>0.3$, then dw$[i]$=3.0 |
| | if SR$[i]>0.4$, then dw$[i]$=4.0 |
| | if SR$[i]>0.5$, then dw$[i]$=5.0 |
| | if SR$[i]>0.6$, then dw$[i]$=6.0 |
| | if SR$[i]>0.7$, then dw$[i]$=7.0 |
| | if SR$[i]>0.8$, then dw$[i]$=8.0 |
| | if SR$[i]>0.9$, then dw$[i]$=9.0 |
| Ramps (NB and SB) | if SR$[i]>0.1$,then dw$[i]$=4.0 |
| | if SR$[i]>0.2$, then dw$[i]$=5.0 |
| | if SR$[i]>0.3$, then dw$[i]$=6.0 |
| | if SR$[i]>0.4$, then dw$[i]$=7.0 |
| | if SR$[i]>0.5$, then dw$[i]$=8.0 |
| | if SR$[i]>0.6$, then dw$[i]$=9.0 |
| | if SR$[i]>0.7$, then dw$[i]$=10.0 |
| | if SR$[i]>0.8$, then dw$[i]$=11.0 |
| | if SR$[i]>0.9$, then dw$[i]$=12.0 |

Note: SR=storage ratio and dw=dynamic weights.

### Vehicle Trajectories Model

Adaptive signal control requires reliable vehicle arrival and departure information at the stop line for optimization. Usually this arrival information is predicted from arrivals detected at the upstream detector line. According to one of the first models for arrival projection (i.e., OPAC), the number of vehicles arriving at and stopped at the stop line during an interval $\Delta t$ is equal to the number of vehicles detected by detectors during an interval $\Delta T=\Delta t$. This holds only for a lane group that is given green signal phase without initial queued vehicles. This projection scheme is not applicable for a red or a green lane group with initial queued vehicles. Due to the queued vehicles present at the stop line, the travel distance becomes the distance between the back of queue and the detector line, which is shorter than the distance between the stop line and detector line. In the case of a signalized diamond interchange, queuing is the major problem involved in the traffic operation and queues particularly exist at the internal links. Hence, it is necessary to explore the vehicle arrivals and discharges at the stop line in a more microscopic way. Depending on the queue length at the stop-line, the detection range $\Delta T$ varies for a particular stop line interval $\Delta t$.

This study uses an example below to illustrate vehicular movements between the detection line and stop line. Assume that vehicle average speed is 48.3 km/h (30 mi/h), average vehicle deceleration rate is 3.4 $\mathrm{m/s^2}$ (11.2 $\mathrm{ft/s^2}$), and queued vehicle spacing is 7 m (23 ft). Other basic parameters for adaptive signal control are obtained using the guidelines provided in the subsection. "Basic Parameters for Implementing DP Algorithm" As shown in Fig. 4, detectors are setback a distance, e.g., 200 m (656 ft), from the stop line. During the detection period, the detectors detect the presence of a vehicle and its speed, which in turn estimates its arrival time at the stop line sometime during the 10 s optimization horizon. Depending on the queue length at the stop line, detection periods are not constant and vary from 16 to 2 s in advance of the corresponding optimization horizon. According to vehicle trajectories model, the detection range is adjusted and each vehicle's detection time is also located. Vehicle trajectories model states that the net increase in queue length shortens the vehicle travel time, therefore, one more (or less) vehicle queued at the stop line increase (or decrease) the detection period by 0.5 s. This simple relation is found by the travel time estimation. After determining the detection range, the number of vehicles detected during that period are projected to the horizon intervals (0–2.5), (2.5–5), (5–7.5), and (7.5–10) at the stop line.



J. Transp. Eng., 2006, 132(8): 629-637

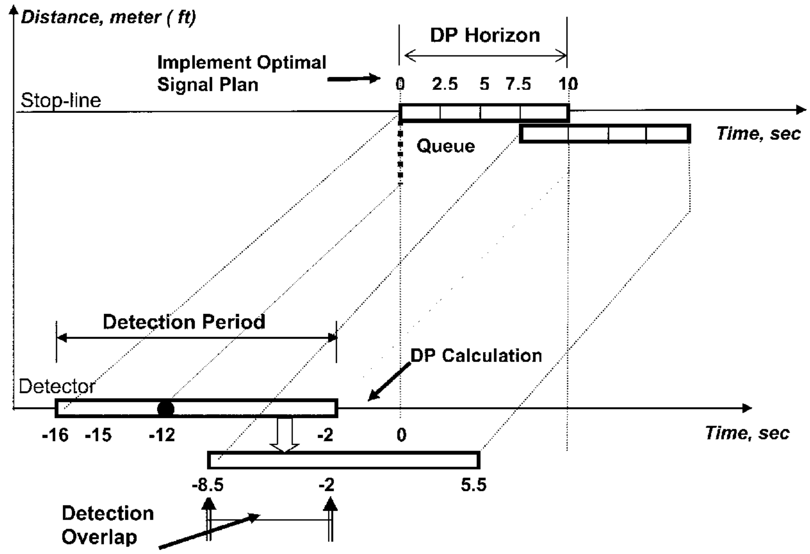


**Fig. 4.** Vehicle trajectories model

To increase the measurement precision and properly implement the majority signal rolling technique (see subsection "Implementation of Optimized Signal Plan"), two consecutive DP optimization horizons are designed to have one interval of 2.5 s overlap. Consequently, there is also an overlap between detection periods. As shown in Fig. 4, two consecutive DP horizons can share the same detection information for up to 6.5 s (−8.5 to ∼−2 s).

## Algorithm Implementation

### Basic Parameters for Implementing DP Algorithm

To implement the DP method developed in the previous section, the algorithm requires the values of the parameters such as the distance of detectors to the stop line, DP horizon, and each interval, and detection range. Those values vary with the field detected data such as average vehicle speed, queue length at the stop line, etc. The basic procedure adopted in this study is to produce a look-up table for vehicle travel time, $T$ (the time taken for a vehicle to travel and stop at the back of queue). Table 3 shows one such look-up table with respect to various queue lengths at the stop line, assuming that vehicle average speed is 48.3 km/h (30 mi/h), average vehicle deceleration rate is 3.4 m/s$^2$ (11.2 ft/s$^2$), and queued vehicle spacing is 7 m (23 ft).

Using the look-up table, the parameters needed to implement the DP algorithm can be determined as follows:

1. Detectors setback at 200 m (656 ft) from the stop line for all lane groups. This setup is selected in order to ensure reliable traffic information of more than 10 s in advance even when there is certain number of queued vehicles present (e.g., 10). The distance also corresponds to the vehicle travel distance with zero queued vehicles at the stop line.
2. DP horizon at stop-line is 10 s and it is divided into four intervals. Each DP interval is 2.5 s. This takes into account the minimum travel time that can be achieved with the predetermined detector setup and detected vehicle speed, and the minimal green time that can be implemented. Minimum travel time represents the minimum time during which traffic information can be obtained in advance.
3. Detection range for vehicle projection is 16 to 2 s in advance of the corresponding DP horizon. This predicts the maximum

**Table 3.** Travel Time at Various Queue Lengths Present

| Number of queued vehicles | $T_1$ Travel time with constant speed (s) | $T_2$ Vehicle braking time (s) | Travel time $T=T_1+T_2$ (s) | Distance from detector to back of queue [m (ft)] |
|---|---|---|---|---|
| 0 | 12.91 | 3.94 | 17 | 200 (656) |
| 1 | 12.38 | 3.94 | 16 | 193 (633) |
| 2 | 11.86 | 3.94 | 16 | 186 (610) |
| 3 | 11.34 | 3.94 | 15 | 179 (587) |
| 4 | 10.82 | 3.94 | 15 | 172 (564) |
| 5 | 10.30 | 3.94 | 14 | 165 (541) |
| 6 | 9.78 | 3.94 | 14 | 158 (518) |
| 7 | 9.26 | 3.94 | 13 | 151 (495) |
| 8 | 8.73 | 3.94 | 13 | 144 (472) |
| 9 | 8.21 | 3.94 | 12 | 137 (449) |
| 10 | 7.69 | 3.94 | 12 | 130 (426) |



J. Transp. Eng., 2006, 132(8): 629-637

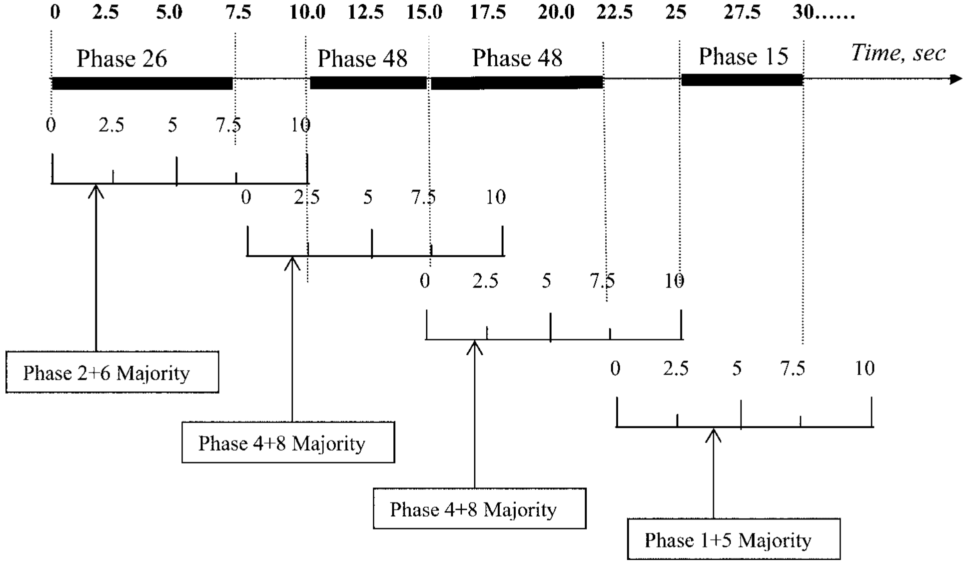


**Fig. 5.** Signal implementation—majority rolling technique

detection range for one optimization horizon. Depending on the queue length at the stop line, the corresponding detection range varies. The boundaries of maximum detection range correspond to the scenarios when there is no queued vehicle at the beginning of the DP horizon and there are maximum queued vehicles (e.g., 10) in the end of the horizon.

### *Detector Setup*

Fig. 2 presents a detectors placement layout for implementing the DP algorithm. A diamond interchange has eight lane groups including NB, SB, EB external, WB external, EB internal, WB internal, EB left turning, and WB left turning. Detectors are located upstream some distance for all movements. These detectors should have the capabilities of measuring the number of vehicles passing through and their speeds.

### *Implementation of Optimized Signal Plan*

To implement the optimized signal plan, a majority rolling technique was developed. This technique can be stated as follows: over a single horizon of 10 s, the signal phase chosen most frequently is implemented for 7.5 s followed by a yellow-and-all-red clearance time 2.5 s; and if two consecutive horizons hold the same majority signal then no clearance time is inserted between them. The majority signal phase is defined as the one option that appears most among four DP optimal solutions for a DP horizon. An illustrative example is given in Fig. 5. In the AIMSUN simulation as shown in the next section, it is assumed that $Y+\mathrm{AR}$ time is 2.5 s. However, depending on interchange geometric characteristics, this change and clearance interval can be adjusted to meet the requirements of the specific site. The specification of the $Y+\mathrm{AR}$ duration does not impact the algorithm itself, but should be considered in its implementation at a specific site.

The majority of signal rolling allows for implementation of sufficient long green and yellow–red intervals by avoiding the frequent switches of the signal phases of small intervals. It also tends to reduce the accumulation of arrival projection errors due to an overlapped interval of 2.5 s. However, this scheme sacrifices some of computations in order to be implemented practically and does not fully represent optimal solutions.

## Simulation Evaluation of Developed Algorithm

The simulation of diamond interchange operation provides a cost-effective way to evaluate the performance of the new DP signal plan strategy before it could be implemented. A procedure for conducting this simulation evaluation is proposed in this research as shown in Fig. 6. First, a microscopic simulation model that can simulate networks and adaptive signal control is selected as an evaluation model. AIMSUN has been selected as the simulation model in this study because its unique application programming interface (API) function named as the GETRAM extension module can be used to implement advanced traffic control applications. Next, the selected model is calibrated using diamond interchange field data. The calibrated diamond interchange is then used to compare the performance of three different optimal signal plans under the same scenarios. Signal plans are obtained from three signal optimization models: PASSER III, TRANSYT-7F, and the DP algorithm. Simulation experiments are conducted on the calibrated models for all comparison cases by using the same and unchanged simulation seed number, driver behavior, and vehicle characteristics to every comparison study. Each case was simulated for 60 min. Finally, this study compares delays of each movement and the entire diamond interchange from these three signal optimization algorithms. The proposed algorithm can also be compared to actuated control. However, this option was not adopted in this study because the settings of actuated control parameters (such as minimum green, maximum green time, etc.) deviate from the PASSER III and TRANSYT-7F settings, and can greatly affect the results. It is recommended that a comparison between the DP algorithm and actuated controls should be conducted in the future. That study should evaluate the impacts of various settings, i.e., minimum green, maximum green, the position of detectors, etc.

The signal plan obtained from the DP algorithm is dynamic and cannot be implemented by directly inputting signal timing to a simulation model. Therefore, in this study the DP algorithm is coded in C++ language and dynamically linked to the AIMSUN, a stochastic microsimulation package, which is used for evaluation and simulation of the developed methodology. AIMSUN simulates a signalized diamond interchange instrumented with



J. Transp. Eng., 2006, 132(8): 629-637

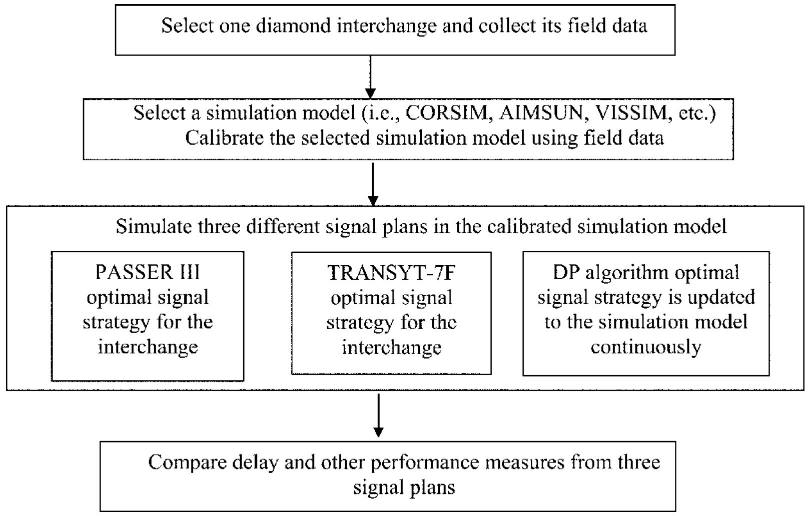


**Fig. 6.** Simulation evaluation procedure

loop detectors that can provides vehicle counts and speeds to the DP algorithm. Based on this, the algorithm calculates the optimal phase sequence and the duration of each phase, and passes them back to AIMSUN, which subsequently controls the interchange in real time. The communication between the AIMSUN and the DP algorithm is achieved through a number of user-defined extension functions that are specified by the GETRAM module.

## Comparisons

The comparison study on the performance of three different optimization methods, i.e., the real-time DP algorithm, PASSER III, and TRANSYT-7F has been conducted using the calibrated interchange under varying demand scenarios. These scenarios include varying demand on ramps, varying demand on arterials, and varying demand on both ramps and arterials. For each demand scenario, the optimized pretimed signal plans from TRANSYT-7F and PASSER III are implemented in AIMSUN and the results are compared to those from the DP algorithm. The signal plans of the DP algorithm are not predetermined but generated every 10 s based on real-time demand situations. The simulation time is 1 h.

The performance measures (i.e., system delay and movement delay) from three different optimization algorithms are compared with respect to each scenario. Table 4 presents the comparisons of the system delay among different optimization methods in terms of varying arterial demand scenarios. All scenarios are balanced demands. Arterials volume increase from 2,000 to 3,300 vehicles/h. The study intends to explore how the interchange performance would be affected by substantially increasing volume on certain lane groups, although some simulation demands are much different from the optimization demand.

When the demand is close to the optimization demand (i.e., arterial 2,000 vehicles/h), each system has similar results. When arterial demand substantially increases, i.e., increase by 65% from 2,000 to 3,300 vehicles/h, the PASSER III and TRANSYT-7F exhibit much larger delays than the DP algorithm does. Both PASSER III and TRANSYT-7F increase the system delay up to 1 min or more. The interchange under the control of the proposed real-time algorithm does not experience delay as much as the off-line systems do. The system delay resulted from fixed weights increases from 21 to 43 s while the delay from dynamic weights remains almost the same with the range between 23 and 29 s. Therefore, compared to fixed weights, using dynamic weights can reduce the system delay up to 32.5% in handling increased arterial demands.

Several other demand scenarios were also designed to study the performance of using dynamic weights in the algorithm. As shown in Table 5, when the demand varies unpredictably every 15 min and is unbalanced, using dynamic weights reduces the system delay by 36–49%, compared to using fixed weights. Dynamic weights have the capability to discharge the increased volume on particular lane group(s) quickly enough while maintaining efficient system operations.

**Table 4.** System Delays (second/vehicle) Comparisons Among Three Optimization Methods

| Number | Scenario | Proposed DP Fixed weight | Proposed DP Dynamic weight | PASSER III | TRANSYT-7F |
|---|---|---|---|---|---|
| 1 | Arterial TH 2,000 vehicles/h | 21 | 26 | 19 | 21 |
| 2 | Arterial TH 2,300 vehicles/h | 26 | 26 | 38 | 33 |
| 3 | Arterial TH 2,500 vehicles/h | 32 | 29 | 51 | 67 |
| 4 | Arterial TH 2,700 vehicles/h | 34 | 23 | 61 | 86 |
| 5 | Arterial TH 3,000 vehicles/h | 40 | 24 | 76 | 82 |
| 6 | Arterial TH 3,300 vehicles/h | 43 | 24 | 81 | 92 |



J. Transp. Eng., 2006, 132(8): 629-637

**Table 5.** System Delay (second/vehicle) of Using Dynamic Weights and Fixed Weights

| Number | Demand scenario | Dynamic weights | Fixed weights |
|---|---|---|---|
| 1 | High EB demand | 30 | 52 |
| 2 | High SB demand | 23 | 36 |
| 3 | High EB and WB demand | 39 | 77 |
| 4 | High EB and SB demand | 29 | 51 |
| 5 | High EB and NB demand | 29 | 48 |

Fang (2004) includes additional comparisons and discussion of the three optimization methods under various demand scenarios. Important findings from those comparisons are as follows:

1. When the field demand is similar to the demand for off-line optimization and it varies little in each 15 min period, all three methods exhibit nearly identical performance in terms of the system delay and each movement delay.
2. The signal plans from the DP algorithm are adjusted automatically in response to real-time demand fluctuations. With varying demands, the off-line systems would cause longer delays either on the arterial or the ramp than the on-line system.

## Conclusions and Recommendations

This research provides a unique method to provide optimal signal control at diamond interchanges based on reliable detected traffic information. The major contributions this research has made are summarized as follows:

1. The adaptive signal optimization problem has been formulated in a decision network with phase sequence and duration as decision variables. The forward recursive DP solution procedure has been developed in the context of real-time optimization. A number of performance measures have been defined and used as the objective functions for the signal optimization based on the detected vehicle information.
2. Vehicle trajectories from the detection until-vehicles have arrived and then stopped or discharged at the stop lines at each movement, have been modeled and an interactive projection scheme has been developed at the microscopic level.
3. The proposed algorithm that incorporates the DP procedure, the projection of the detected vehicles, and the arrival–discharge dynamics has been implemented in C++. It has also been dynamically linked to AIMSUN, which has made possible the real-time simulation of diamond interchanges using the optimized signals under different demand scenarios.
4. The developed real-time adaptive signal algorithm can optimize both phase sequence and phase duration by minimizing a user prespecified performance measure over a finite horizon that rolls forward.
5. The real-time DP signal algorithm is superior to PASSER III and TRANSYT-7F which output pretimed optimal results in handling demand fluctuations.
6. The dynamic weighted DP algorithm is appropriate to be applied in special events or incidents when high demands are unexpected and varying.

Upon the completion of the study, the following have been identified and recommended for future research:

1. The current decision network formulated for the signal optimization of a diamond interchange is based on a three-phase ring structure. This structure should be extended to include every possible combination of phases for diamond interchange traffic operations. The absence of these phases is acknowledged in the present algorithm by assigning weights on ramp approaches in order to discharge ramp vehicles efficiently.
2. The developed algorithm requires a detector set upstream of the stop line on each lane group of a diamond interchange. When it is not possible or practical to place detectors far enough from the intersection to provide reliable vehicular information, for example at tight interchange configurations, the use of predicted traffic data are suggested, and the traffic flows at the internal link are to be obtained from detected traffic information at external approaches.

## References


Charles, P. (2001). "Time for a change." *Traffic Technol. Int.*, June/July, 68–71.

Fambro, D. B., Chaudhary, N. A., Bonneson, J. A., Messer, C. J., and Arabie, L. L. (1991). PASSER III—90 users manual and application guide, Texas Transportation Institute, College Station, Tex.

Fang, C. F. (2004). "Optimal adaptive signal control for diamond interchanges using dynamic programming." Ph.D. thesis, Pennsylvania State Univ., University Park, Pa.

Gartner, N. H. (1983). "OPAC: A demand-responsive strategy for traffic signal control." *Transportation Research Record. 906*, Transportation Research Board, Washington, D.C., 75–81.

Gartner, N. H., and Pooran, F. J. (2002). "Implementation and field testing of the OPAC adaptive control strategy in RT-TRACS." *Proc., 81st Annual TRB Meeting*, Transportation Research Board, Washington, D.C.

Hastings, N. A. J. (1973). *Dynamic programming with management applications*, Butterworths, London.

Head, K. L. (1995). "Event-based short-term traffic flow prediction model." *Transportation Research Record. 1510*, Transportation Research Board, Washington, D.C., 45–52.

Hunt, P. B., et al. (1982). "The SCOOT on-line-traffic-signal optimization technique." *Traffic Eng. Control*, 23(4), 190–192.

Luk, J. Y. K. (1984). "Two traffic responsive area traffic control methods: SCATS and SCOOT." *Traffic Eng. Control*, 30, 14–20.

Messer, C. J., and Bonneson, J. A. (1997). "NCHRP project 3-47 final report: Capacity analysis of interchange ramp terminal." *Rep., NCHRP Web Document 12*, Transportation Research Board, National Research Council, Washington, D.C., ⟨http://www.nap.edu/readingroom⟩

Messer, C. J., Fambro, D. B., and Richards, S. H. (1997). "Optimization of pretimed signalized diamond interchanges." *Transportation Research Record. 644*, Transportation Research Board, Washington, D.C., 78–84.

Miller, A. J. (1963). "A computer control system for traffic networks." 200–220.

Mirchandani, P., and Head, L. (2001). "A real-time traffic signal control system: Architecture, algorithms and analysis." *Transp. Res., Part C: Emerg. Technol.*, 9, 415–432.

Mirchandani, P., and Lucas, D. E. (2001) "RHODES-ITMS TEMPE FIELD test project: Implementation and field testing of RHODES, a real-time traffic adaptive control system." *FHWA-AZ01-447*, Federal Highway Administration, Washington, D.C.

Sen, U., and Head, K. L. (1997). "Controlled optimization of phases at an intersection." *Transp. Sci.*, 31(1), 5–17.

Transportation Research Board, (2000). *Highway capacity manual*, National Research Council, Transportation Research Board, Washington, D.C.




J. Transp. Eng., 2006, 132(8): 629-637