

UC Berkeley

Research Reports

Title

Vehicle-Actuated Control of a Diamond Interchange

Permalink

<https://escholarship.org/uc/item/31p1s9df>

Author

Newell, Gordon F.

Publication Date

1998-07-01

Institute of Transportation Studies
University of California at Berkeley

Vehicle-Actuated Control of a Diamond Interchange

Gordon F. Newell

RESEARCH REPORT
UCB-ITS-RR-98-3

July 1998
ISSN 0192 4095

Vehicle-Actuated Control of a Diamond Interchange

G.F. Newell

Institute of Transportation Studies
University of California, Berkeley 94720

Abstract

We present here a vehicle-actuated traffic signal control scheme for a simple geometry of a diamond interchange. The scheme is guaranteed not to have blockage, to adjust automatically to changing flows and operate on a shorter mean cycle time than most fixed-time plans, at least during times of heavy traffic.

There have been many research reports on how to signalize a diamond interchange and traffic control manuals [1] contain various recipes for doing so. The objective in these strategies seems to be that (most) vehicles which leave one intersection should not be stopped at the second intersection. If one succeeds in doing this, however, it imposes very severe constraints on the duration of the various signal phases and the cycle time. Different schemes are used depending on the flows and fractions of turning vehicles, but, under heavy traffic, they all suffer from the possibility that queues may accumulate between the signals and block vehicles from entering the interchange or turn bays.

For light traffic there may be no delay to a vehicle at its second intersection but the delay at its first intersection may be larger than necessary because the cycle time is larger than necessary to accommodate the flows. Some signals have traffic responsive controllers but, except for very light traffic (serving perhaps only one vehicle per phase), they typically operate with some pretimed plan.

We present here a vehicle-actuated control scheme which will automatically adjust to any changes in the approach flows and also guarantee that queues will not accumulate between the signals. We consider only one simple type of geometry as illustrated schematically in Fig. 1a, but one could presumably modify the scheme to deal with other geometries. Typically there will be at least one extra highway lane between the signals for turn bays. We might for illustrative purposes suppose that there is a single lane for through traffic in each direction and a single lane for the two left turn bays.

We will assume here that right turning vehicles have a negligible effect on the signal timing. In Fig 1a the critical traffic streams are labeled 1, 2, 3, 4 at intersection I, 1', 2', 3', 4' at intersection II. Movement 1 (or 1') is a left turn to a freeway on-ramp, 2 (or 2') is approaching the interchange on the through road, 3 (or 3') is the through traffic from within the interchange, and 4 (or 4') is the left turn traffic from the freeway off-ramp.

Each signal will have three phases. Phase A (or A') serves the freeway off-ramp, stream 4 (or 4'), phase B (or B') can serve movements 3 and 2 (3' and 2') simultaneously, and phase C (or C') can serve movements 3 and 1 (3' and 1') simultaneously. Vehicle entering the interchange in streams 2' (2) will split with some joining streams 3' (3) and some entering the turn bay joining stream 1' (1). We assume, however, that vehicles coming off the freeway, stream 4 (4'), are not likely to make a U-turn and go back onto the freeway in stream 1' (1). Nearly all vehicles from stream 4 (4') will join stream 3' (3).

Unlike the conventional signal timing schemes which are directed toward trying to provide "through bands" for certain traffic movements with no queues between the intersections, the scheme proposed here will intentionally generate temporary queues so that, when it is time to serve some traffic stream, it will be served at the maximum rate, the saturation flow for that movement. The emphasis will be on timing the signals so that the queues for one traffic movement will not block any other traffic movements. In particular a queue in the through lane should not block vehicles which want to enter the turn bay, and queues in the turn bay should not spill over into the through lane to block the through vehicles.

There are several other features which we would like the signal timing scheme to have:

1. The signal timing should automatically adjust to changes in the arriving flow pattern.
2. If the interchange should become oversaturated, any residual queues should be stored on the approaches. The highway segment between the signals should not fill so as to block some traffic streams from entering the interchange.
3. Any queue in the highway section between the signals should vanish at the end of the signal phase in which it is being served. There should be no internal residual queues carried over from one cycle to the next. If there were a residual queue in one cycle, there could be one also in subsequent cycles which might eventually cause a blockage.

The Strategy

Figure 1b shows a time sequence of phases which satisfies the above conditions.

We start with phase A at intersection I which generates mostly through traffic at intersection II. This phase will end as soon as the queue for stream 4 vanishes or this phase has lasted some prespecified maximum time.

When vehicles from stream 4 reach the second intersection, they may join a queue in the through lane, but we want to be sure that these vehicles will not block the entrance to the turn bay for subsequent arrivals from stream 2. One way to do this is to make sure that when phase A' starts (and both streams 1' and 3' are blocked) no vehicles are arriving from stream 4 and there is no queue in either the turn bay or the through lane. This means that phase A' should follow phase C' and phase C' should last until (1) the last vehicle from phase A could reach intersection II a free flow trip time τ after the end of phase A, (2) the queue in the through lane vanishes and (3) the queue in the turn bay vanishes (whichever occurs last).

Condition (1) provides a somewhat flexible coupling between the two signals. If, without this condition, the average cycle time at II would be shorter than the average cycle time at I, then this condition would likely determine the end of phase C'. Otherwise the two signals will likely be linked by a corresponding condition later at the end of phase C, at a time no earlier than a time τ after the completion of phase A'.

If both queues in phase C' vanish before a time τ after the end of phase A, there will still be vehicles arriving from phase A until approximately a time τ after the end of phase A. Indeed they may be arriving at a rate comparable with the saturation flow for stream 4. After these vehicles have been served, however, there should be a significant drop in flow because there is a yellow signal after phase A and vehicles from phase B would not likely reach the second intersection before phase C' ends.

In this case there is some time "wasted". If the queue in the turn bay vanishes before the end of phase C', any newly arriving vehicles are coming from stream 4 and are not likely to provide any new vehicles for the turn bay. For a (short) time the turn signal will be green, but there is no one there.

If phase C' continues beyond a time τ after the end of phase A in order to serve the queues in streams 1' or 3' (more likely 1'), then some vehicles from phase B, stream 2, might reach the queues at the second intersection before phase C' ends. If they can join the queue, they will be served in phase C' but phase C' should terminate as soon as the flows in both streams 1' and 3' drops below their saturation flows. This will happen because the arrival rates in both streams 1' and 3' should be well below the saturation flows. Note that there is no possibility that the turn bay will be blocked, because both queues are dissipating in phase C'.

Phase C' is followed by A' after an appropriate yellow or simultaneous red to clear the intersection. Like phase A, A' ends when the queue in stream 4' vanishes or it has reached some prespecified maximum.

Phase A (A') are followed by phase B(B') and then C(C'). Phase B(B') will end when the queue in stream 2($2'$) vanishes or this phase has lasted some prespecified maximum time. The prespecified maximum will be chosen so that the queue generated by the vehicles from stream 2($2'$) into stream 1' (1) is not likely to overflow the turn bay and thus cause a possible blockage in phase B' (B). If the phase runs to the maximum, this means that the interchange is (temporarily) oversaturated because one is leaving a residual queue in stream 2($2'$).

One cannot know precisely how many vehicles from stream 2($2'$) will enter the turn bay and how many are through vehicles until they actually reach the turn bay. But by the time they fill the turn bay it is already too late to terminate phase B(B'). The maximum time for phase B(B') could be chosen based upon what happened in previous cycles or on other days when the interchange was oversaturated. Alternatively one could terminate phase B(B') if the number of vehicles in the turn bay reaches some prespecified number or occupies some fraction of the length of the turn bay (maybe 2/3). It is the storage capacity of the turn bay which will dictate the capacity of the interchange.

Finally the cycle at intersection I ends with phase C which, like phase C' , ends when the queues in streams 1 and 3 both vanish and it is at least a time τ after the end of the phase A' .

This strategy satisfies all the conditions listed in the previous section. Except for the coupling at the ends of phase C' and C, the two signals operate independently. The preset maximum times will restrict the number of vehicles which can enter the interchange, but any vehicles which enter will be served within the next cycle regardless of which queue they join (condition 3). The lengths of the various phases automatically adjust to changes in the arriving flow pattern (condition 1).

If the signal becomes oversaturated, the preset maximum times on phases A, A' , B and B' will keep any residual queues on the approaches (condition 2).

The Mean Cycle Time

We next wish to see what restrictions there may be on the flows which can be accommodated and the cycle time.

Suppose the flow that is served in stream j , j' is q_j, q'_j ($j = 1, 2, 3, 4$). These are not independent since they must satisfy

$$\begin{aligned} q_4 + q_2 &= q'_1 + q'_3, \\ q'_4 + q'_2 &= q_1 + q_3 \end{aligned}$$

so there are only six independent flow parameters.

Since we assume that a negligible fraction of stream 4 will join stream 1', it may be advantageous to define

$\kappa \kappa' =$ fraction of vehicles in stream 2, 2' which join stream 1', 1

so that

$$\begin{aligned} q_1' &= \kappa q_2, \quad q_1 = \kappa' q_2' \\ q_3' &= q_4 + (1 - \kappa q_2), \quad q_3 = q_4' + (1 - \kappa') q_2' \end{aligned} \quad (1)$$

These are still six independent parameters, but they are now

$$q_2, q_4, q_2', q_4' \quad \text{and} \quad \kappa, \kappa'$$

the four input flows to the interchange and the two turning fractions.

If the signal is oversaturated one or more of the flows q_2, q_4, q_2', q_4' may be constrained to what can actually be served.

From Fig. 1b we see that the time needed for phases A plus B is the largest of

- a. the time needed to serve stream 3
- b. the time needed to serve streams 2 plus 1
- c. 2τ plus the time needed to serve stream 4' plus any extra time needed to finish serving streams 1' and 3' if phase C' last longer than a time τ after phase A.

The cycle time to complete phases A plus B plus C is this plus the time needed to serve stream 4.

If the saturation flows for streams j, j' are s_j, s_j' and the average cycle time is T , the average time per cycle needed to serve streams j, j' are $Tq_j/s_j, Tq_j'/s_j'$. There will also be some effective lost times L_{jk}, L'_{jk} in switching from serving stream j to k, j' to k' .

Which of the times a, b, c above is largest may change from one cycle to the next. The average of the largest of these times, however, must be at least as large as the largest of the averages. From this we conclude that

$$L_{43} + L_{34} + Tq_3/s_3 + Tq_4/s_4 \quad (2a)$$

$$T \geq L_{42} + L_{21} + L_{14} + Tq_2/s_2 + Tq_1/s_1 + Tq_4/s_4 \quad (2b)$$

$$2\tau + L'_{14} + L_{14} + Tq_4'/s_4' + Tq_4/s_4 \quad (2c)$$

There are also corresponding inequalities with the primed and unprimed variables interchanged. The last condition above is already symmetric to this interchange but we must also have

$$L'_{43} + L'_{34} + Tq'_3/s'_3 + Tq'_4/s'_4 \quad (2d)$$

$T \geq$

$$L'_{42} + L'_{21} + L'_{14} + Tq'_2/s'_2 + Tq'_1/s'_1 + Tq'_4/s'_4 \quad . \quad (2e)$$

These conditions, expressed in terms of the variables (1) imply that

$$\frac{L_{43} + L_{34}}{1 - q'_4/s_3 - (1 - \kappa')q'_2/s_3 - q_4/s_4} \quad (3a)$$

$$\frac{L_{42} + L_{21} + L_{34}}{1 - \kappa'q'_2/s_1 - q_2/s_2 - q_4/s_4} \quad (3b)$$

$$T \geq \frac{2\tau + L_{14} + L_{41}}{1 - q_4/s_4 - q'_4/s'_4} \quad (3c)$$

$$\frac{L'_{43} + L'_{34}}{1 - q_4/s'_3 - (1 - \kappa)q_2/s'_3 - q'_4/s'_4} \quad (3d)$$

$$\frac{L'_{42} + L'_{21} + L'_{34}}{1 - \kappa q_2/s'_1 - q'_2/s'_2 - q'_4/s'_4} \quad . \quad (3e)$$

Of course, a necessary condition that the flows can be served is that all the denominators on the right hand side are positive, but there will be more severe restrictions on the flows than this if we impose maximum times on phases A, A', B, and B'.

If the traffic stream which dictates the termination of phases C or C' change from one cycle to the next due to fluctuations in the actual number of vehicles arriving in each cycle, then the mean cycle time could be (somewhat) larger than the largest of the expressions on the right hand side of (3). If, however, the same traffic streams dictate the end of the phases C, C' every cycle, the mean cycle time T will be the largest of the quantities on the right hand side.

For light traffic the condition (3c) is likely to determine the T because, for typical distances between the signals, 2τ is likely to be at least 20 sec. whereas the L_{jk} are likely to be about 5 sec.

The numerator in (3c), therefore, might be about 30 secs, whereas the numerators in (3a) or (3d) may be only 10 sec. and those in (3b) or (3c) about 15 sec. Furthermore one might impose a minimum time on phases A, A'. If so the signal would operate on essentially a fixed (minimum) cycle time (of about 40 sec), but it may not be the most desirable fixed time plan.

The present scheme was devised primarily to provide a saturation flow from the turn bays and to prevent queues from blocking any movements, but for light traffic these are not important issues. If the present scheme is not the most desirable for light flow, it is at least acceptable.

As the flows increase the largest expression on the right hand side of (3) will eventually become the one with the smallest denominator, and this is not likely to be (3c). If $s'_3 = s'_4, s'_3 = s_4$ the denominators in (3a), (3d) would certainly be smaller than in (3c) for $q'_2, q_2 > 0$. Of these two, (3a) or (3e), the dominate one is that which carries the largest through traffic, $(1 - \kappa')q'_2$ or $(1 - \kappa)q_2$. If, however, the traffic from one of the turn bays dictates the cycle time, then the largest term on the right-hand side of (3) will be (3b) or (3e).

Which traffic movements govern the cycle time is likely to change throughout the day. In the off-peak, (3c) will likely dominate. If during the morning peak, traffic turning onto the freeway determines the cycle time, (3b) or (3c), then in the evening peak the traffic leaving the freeway is likely to dominate, (3a) or (3d).

For sufficiently heavy demand, the cycle time and the maximum service rates will be constrained by the storage capacity of the turn bays, and a queue will form on one or more of the approaches. If

$$n_1, n'_1 = \text{average number of vehicles which can be stored in the turn bay for stream 1, } 1',$$

$$G_2, G'_2 = \text{maximum times for phases B, } B',$$

then the average number of vehicles entering the turn bays will be $\kappa s_2 G_2, \kappa' s'_2 G'_2$. The G_2, G'_2 must be constrained so that

$$\kappa s_2 G_2 < n_1 \quad \text{and} \quad \kappa' s'_2 G'_2 < n_1, \quad (4)$$

which in turn will restrict the flow q_2, q'_2 to

$$q_2 < \frac{s_2 G_2}{T} = \frac{n_1}{\kappa T}, \quad q'_2 < \frac{s'_2 G'_2}{T} = \frac{n_1}{\kappa' T}. \quad (5)$$

If the arriving flows exceed these values, any excess demand must wait in a queue.

This, in itself, does not restrict the cycle time. It only restricts the phase times G_2 and G'_2 for serving streams q_2, q'_2 . If there is no maximum time on the phases A and/or A', these

phases can run until the queue vanishes in stream 4 or 4'. If the flows from the freeway q_4 and/or q'_4 should increase, this would extend the phases A and/or A'. But phases C' or C will then extend at least until the queues in streams 3' or 3 have been served. Presumably phases C' or C could accommodate whatever traffic was supplied by phases A or A', at the expense of increasing the total cycle time and, therefore, further limiting the flows q_2 or q'_2 in (5).

If one also imposes maximum times G_4, G'_4 on phases A, A', this would provide an upper limit on the mean cycle time. If the phases A or A' ran to the maximum, then residual queues will form in streams 4 or 4'. By adjusting the times G_2, G_4 or G'_2, G'_4 one can partition the available capacity in any way one chooses between streams q_2, q_4 or q'_2, q'_4 . If the interchange is oversaturated, queues will form someplace. One should store excess vehicles where they will do the least damage to other traffic streams not passing through the interchange.

If the turn bays can store maybe 8 or more vehicles, the size of the turn bays may not be an issue. The above constraints on the maximum times may lead to a cycle time which is unacceptably long. The maximum times may be set at some acceptable values to give a cycle time no larger than perhaps 90 seconds.

Implementation

To implement the above strategy will require vehicle detector configurations different from those in common use now and some redesign of the controller circuitry. The termination of each signal phase depends upon knowledge of when some "queue vanishes". The "queue", however, consists of discrete vehicles and vehicle detectors can only record the passing of individual vehicles or the presence of vehicles over some loop detectors.

While a queue is discharging we expect the average headway, $1/s_j$, to be about 2 seconds (for a single lane), but to be significantly larger than this after the queue vanishes ($1/q_j$ for streams 2, 4, 2' 4'). The most efficient scheme for detecting the end of the queue, however, is to detect a certain spacing larger than some threshold. If one has a long loop detector or a series of small loop detectors in a single lane (as illustrated in figure 1a), then while there is a queue in that lane, stationary or moving, there will be a vehicle over (one of) the detector(s). After the queue has discharged, the next approaching vehicle (if any) should be traveling at some typical running speed but at a spacing larger than for queued vehicles. If the length of the detector configuration is such that it would take about 3-1/2 seconds for a vehicle to traverse the detector at some typical speed, then, at the moment the detector becomes unoccupied, we know that a vehicle has just left the front end of the detector and the next vehicle will be at least 3-1/2 seconds later. Presumably the last "queued" vehicle has just left the detector.

There is a risk that one could accidentally have a 3-1/2 second headway while the queue is discharging (because some slow vehicle cannot keep up with the vehicles ahead), but this cannot happen until the signal has already served any vehicles which were stopped on the detector at the start of the phase.

It would be advantageous to place the front end of the detector about 30 feet or so back from the stop line. If, at the moment the detector becomes unoccupied, the signal will switch to yellow, then there will be a vehicle 30 feet from the stop line. Since the vehicle cannot stop in a distance of 30 feet, it will proceed through the yellow signal. This will reduce the effective lost times L_{jk} because some of the yellow time is utilized.

Phases A, B, A', B' terminate when the queue vanishes in lanes 4, 2, 4', 2' respectively. To terminate phase C(C') one should have three switches in series. One switch is released a time τ after the termination of phase A' (A). Each of the other two switches is released when a detector in streams 1, 3(1', 3') is first unoccupied (not to be reactivated if the detector is occupied again before the signal switches). The signal changes when all switches are released.

If there should be two highway lanes serving any traffic stream, one would expect the queue to vanish in the two lanes at nearly the same time since newly arriving vehicles would likely join the shorter queue. It would, however, be advantageous to have detectors as described above in each of the lanes. To minimize the time lost in switching the signal, one would switch the signal when either detector is unoccupied. This might leave a small residual queue in the lane whose detector is still occupied but this residual queue will not grow from one cycle to the next because newly arriving will tend to equalize the queues in the two lanes.

Conclusion

The timing scheme described here for the simple geometry of Fig 1. has not been field tested, but it is certain to work more efficiently under heavy traffic than existing systems with a fixed time plan or vehicle-actuated controllers with detectors in the wrong places. Once it has been tested on a simple geometry one can try to devise scheme for more complex geometries (with frontage roads, nearby intersections, etc.)

[1] "Traffic Control Systems Handbook", FHWA-1P-85-1, p.3.51 and references therein.