



ELSEVIER

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Self-organizing traffic signals using secondary extension and dynamic coordination

Burak Cesme^{a,*}, Peter G. Furth^b^a AECOM, 2101 Wilson Boulevard, Suite 800, Arlington, VA 22201, United States^b Department of Civil and Environmental Engineering, Northeastern University, United States

ARTICLE INFO

Article history:

Received 19 February 2014

Received in revised form 4 June 2014

Accepted 4 August 2014

Keywords:

Self-organizing signals

Actuated control

Secondary extension

Dynamic coordination

Transit signal priority

Arterial coordination

ABSTRACT

Actuated traffic signal control logic has many advantages because of its responsiveness to traffic demands, short cycles, effective use of capacity leading to and recovering from oversaturation, and amenability to aggressive transit priority. Its main drawback has been its inability to provide good progression along arterials. However, the traditional way of providing progression along arterials, coordinated-actuated control with a common, fixed cycle length, has many drawbacks stemming from its long cycle lengths, inflexibility in recovering from priority interruptions, and ineffective use of capacity during periods of oversaturation. This research explores a new paradigm for traffic signal control, “self-organizing signals,” based on local actuated control but with some additional rules that create coordination mechanisms. The primary new rules proposed are for secondary extensions, in which the green may be held to serve an imminently arriving platoon, and dynamic coordination, in which small groups of closely spaced signals communicate with one another to cycle synchronously with the group’s critical intersection. Simulation tests in VISSIM performed on arterial corridors in Massachusetts and Arizona show overall delay reductions of up to 14% compared to an optimized coordinated-actuated scheme where there is no transit priority, and more than 30% in scenarios with temporary oversaturation. Tests also show that with self-organizing control, transit signal priority can be more effective than with coordinated-actuated control, reducing transit delay by about 60%, or 12 to 14 s per intersection with little impact on traffic delay.

© 2014 Elsevier Ltd. All rights reserved.

1. Common traffic signal control paradigms

In practice, traffic signal control mainly follows two paradigms (Koonce et al., 2008). One is actuated control, in which the sequence by which traffic streams are served is predetermined, but cycle length and the time devoted to each traffic stream varies on a cycle-by-cycle basis to demand. Each traffic stream is served in turn (unless skipped for lack of demand) until a detector senses a gap in traffic (“gaps out”), whereupon it yields to the next conflicting traffic stream, subject to minimum and maximum greens. The other is coordinated control in which all of the signals along an arterial (or in a network) have a common, fixed cycle length, with offsets chosen to allow certain traffic movements to progress with little delay through multiple signals. With fixed-time coordinated control, all signal timing parameters – sequence, cycle length, splits, and offsets – are fixed. In the dominant variation of coordinated control applied in the U.S., “coordinated-actuated” control, splits are nominal,

* Corresponding author. Tel.: +1 703 340 3119.

E-mail addresses: burak.cesme@aecom.com (B. Cesme), p.furth@neu.edu (P.G. Furth).

and minor traffic streams that gap out before the nominal end of their phase will yield control early. The cycle length is held fixed, however, with slack time due to gap-out typically going to the “coordinated” phase, usually the arterial through movement. A coordinated phase may begin before but never after its nominal start, and always ends at its nominal end.

Actuated control is applied mainly at isolated intersections, while coordinated control tends to be favored along arterials and in networks where signal spacing is short enough to prevent wide platoon dispersion. The Manual on Uniform Traffic Control Devices recommends that signals within 0.5 mile be coordinated under a common cycle length ([Federal Highway Administration, 2009](#)).

1.1. Advantages of actuated versus coordinated control

Compared to coordinated control with its fixed cycle length, actuated control has many desirable properties. By matching phase lengths in every cycle to the time needed to dissipate the queues, actuated control prevents overflow (a remaining queue when a signal turns red) during undersaturated conditions, and otherwise keeps the cycle length as short as possible. This combination reduces delay to all users unable to take advantage of progression, including pedestrians, transit, turning traffic, and minor-street cross traffic. Where all traffic arrives randomly, actuated control logic has been shown to achieve near-minimal delay for motor traffic ([Newell and Osuma, 1969](#)). In contrast, where cycle length is fixed, minimizing delay requires inserting slack time into the cycle in order to limit the probability of overflow due to random traffic fluctuations. The remaining probability of overflow and the increased cycle length both increase delay to all users except those on the coordinated movement.

The requirement of a common cycle length can drastically increase cycle length at many of the intersections within a coordinated scheme. The cycle length needed by any given intersection to serve its demand varies depending on several factors including the number of mutually conflicting phases, traffic volumes and distribution, and lane availability; it is given by the formula:

$$C_{\text{needed}} = \frac{\sum_{\text{critical}} L_i}{1 - \sum_{\text{critical}} \frac{v_i}{s_i X_{\text{target},i}}} \quad (1)$$

where C_{needed} = an intersection's needed or natural cycle length, sums are over critical movements only, v_i and s_i are movement i 's arrival rate and saturation flow rate, L_i = movement i 's lost time, and $X_{\text{target},i}$ = movement i 's target degree of saturation ([Koonce et al., 2008](#)). First, while $X_{\text{target},i} = 1$ with actuated control, with coordinated control $X_{\text{target},i}$ must be less than 1 in order to limit the chance of overflow, making each intersection's needed cycle length longer. But more importantly, with coordinated control, every intersection is forced to have the cycle length of the critical intersection. Thus, for example, with actuated control the critical intersection may cycle (on average) every 90 s, while another tends to cycle every 75 s and another every 60 s; with coordinated control, the critical intersection is likely to need a cycle of 100 s, and all of the intersections will then be forced to cycle at 100 s. The delay experienced by pedestrians, transit vehicles, and uncoordinated traffic movements is roughly proportional to cycle length, and will therefore be far greater when the cycle length is 100 s instead of 60 s. For example, average delay to pedestrians crossing the main street is roughly half the cycle length, and so increasing cycle length from 60 to 100 s in order to bring an intersection into coordination will increase its pedestrian delay by about 20 s.

The large amount of slack time forced by coordination into cycles at many intersections gives rise to long periods of unsaturated green, that is, periods in which the arterial through movement has green even though it has little or no flow. These long periods of unsaturated green promote speeding on the main line and risky crossings by pedestrians and turning motorists. Actuated control, in contrast, minimizes unsaturated green, because its operating principle is to switch control as soon as the queue has discharged. Many traffic engineers consider this a safety advantage of actuated control.

When recovering from temporary oversaturation, actuated control will make full use of capacity, ending the green on phases whose queue has dissipated and switching to phases that still have a queue. Coordinated control, on the other hand, will allow the arterial phase to remain green in the absence of a queue while queues persist on minor phases, thus wasting capacity during periods leading up to and recovering from oversaturation. In addition, offsets chosen for undersaturated conditions can lead to queue interactions that limit capacity during oversaturated conditions. The poor performance of traditional coordinated control with respect to oversaturation is well known ([Abu-Lebdeh and Benekohal, 2000](#)).

Finally, actuated control has compensation mechanisms that make it amenable to aggressive transit signal priority. If a priority interruption cuts a phase short and creates longer-than-usual queue, that phase's green will automatically adjust its length in the next cycle in order to dissipate its queue (subject to maximum green). A compensation mechanism like this is a critical aspect of making traffic signal control well suited to transit priority ([Furth et al., 2010](#)). Actuated operation with aggressive priority is commonly used in many European cities including Zurich and The Hague, reducing delay for trams and some bus lines to nearly zero ([Nash, 2001](#)). In contrast, coordinated control lacks a mechanism to compensate any phase whose green time was cut short except the coordinated phase, and so queues that result from priority interruptions can persist for several cycles, resulting in long delays. In order to avoid these long delays, it is common for priority to be sharply curtailed by constraints such as limiting green extension lengths, disabling priority for one or more cycles after a priority action is taken, and disabling priority when detectors on a cross street indicate a high level of occupancy. These limitations can seriously impair the effectiveness of transit priority. Published transit priority studies from the U.S. indicates that in

many cases, average delay reduction to transit is 3 s per intersection or less (Dion et al., 2004; Kloos et al., 1994). A primary motivation for this research was our belief that the limited effectiveness of transit signal priority as applied in America stems partly from the inflexibility of the coordinated control paradigm within which it is forced to operate.

1.2. Progression benefits of coordinated control

Coordinated control offers one powerful advantage over actuated control on arterials with frequent signalized intersections: offsets can be chosen to provide a green wave in one arterial direction, giving the through traffic in that direction near-zero delay. That makes this scheme especially suitable for one-way arterials. On two-way arterials, good progression can be provided in both directions if intersection spacing is close to ideal, which means the travel time between intersections equals a multiple of half the cycle length. One study estimates that coordination can reduce delays and stops by between 10% and 40%, depending on the prior method of signal control, traffic flows, and intersection spacing (Robertson and Bretherton, 1991). Where intersections are closely spaced, coordination can be used to prevent queues from an intersection from starving a downstream intersection, or from allowing its queue to block an upstream intersection.

However, realistic intersection spacing and traffic patterns often make it such that the coordination benefits of coordinated control fall far short of the ideal of everybody enjoying a green wave. The irregular and non-ideal intersection spacing prevalent on many arterials makes it such that good two-way coordination is simply not achievable; progression will either be good for one direction and poor for the other, or a compromise solution will be found in which stops are frequent in both directions. In urban settings, a large fraction of arterial traffic does not continue through, but turns on or off at some point, reducing the fraction of traffic that can benefit from progression and also reducing platoon density, which undermines the efficiency of control aiming to provide green waves.

1.3. Adaptive control

An additional weakness of coordinated control is that timing parameters are optimized for traffic volumes collected in the past, usually for only a few selected periods of the day, and can therefore be a poor fit with traffic conditions for much of the day. The most commonly used forms of adaptive control, including SCOOT, SCATS, and ACS-Lite, overcome this particular weakness by frequently updating cycle length, splits, and offsets based on recent traffic measurements (NCHRP, 2010). However, in other respects those adaptive control methods follow the coordinated control paradigm with all of the weaknesses related to fixed cycle lengths. In addition, when facing increased traffic volume leading to oversaturation, most adaptive control methods with fixed cycle length increase cycle length, hoping that reducing the lost time per cycle associated with switching will increase throughput (NCHRP, 2010). In practice, however, these longer cycles often fail to yield greater capacity and sometimes actually degrade performance (Campbell and Skabardonis, 2014), because the longer queues associated with longer cycles can block auxiliary lanes and thereby reduce capacity (Denney et al., 2009). Moreover, when aggressive TSP was applied, adaptive control methods with fixed cycle length lack the inherent flexibility to recover from priority interruptions as they allow only small transitions in cycle length and split in every timing update.

Adaptive control methods free of any cycle length such as RHODES and UTOPIA distinguish themselves from other adaptive methods by not imposing coordination externally, but rather allowing it to emerge from the application of optimization rules using information shared between neighboring intersections. The success of an optimization approach relies on an ability to predict traffic arrivals over the decision horizon, which is difficult in urban networks because arrival patterns are not exogenous but are themselves an output of the decision process made at neighboring intersections. PAMSCOD (He et al., 2012) is a new adaptive method in this class, using a scheduling framework in which the decision is whether to serve imminently arriving platoons and transit vehicles in the first or second cycle at both the first and second intersection they reach. The intent is to see coordination emerge from repeatedly solving local scheduling problems that considering two successive intersections. Reported tests indicate significantly less delay for both buses and traffic under many conditions, including during temporary oversaturation, but the computational burden is impractical under high demand.

2. Self-organizing control

A system is described as self-organizing when local elements interact with each other in order to achieve dynamically a global function or behavior. Self-organizing is a process whereby the pattern at the global level emerges solely from the interactions among the lower level or local components of the system (Camazine et al., 2001). Self-organizing control would seem to be ideal for traffic signal control, where traffic demands vary both randomly and systematically, especially where transit priority interruptions are contemplated. It has attracted the attention of several researchers.

Gershenson (2005) and Gershenson and Rosenblueth (2012), discussed in a later section, make the strongest case for self-organizing traffic signal control. Xie et al. (2011, 2012) propose a self-scheduling approach, in which each intersection is controlled by a local agent operating with a limited horizon. The proposed logic aimed at deciding whether or not to extend a green phase when there is no traffic in order to serve future traffic (i.e., platoons and anticipated queues), with the goal of establishing green waves between successive signals. The control logic considered two policies: (1) "platoon-based extension" in which the current green phase is extended to serve an arriving platoon based on the flow rate of a platoon, and

(2) “platoon-based squeezing” in which a green phase can be truncated in order to provide a green wave to a platoon expected to arrive on an approach with a red signal. Their results indicated that the proposed logic outperforms pre-timed traffic control as well as an adaptive strategy with no coordination between neighbors. However, the control logic was tested on one-way streets with no turning vehicles at most intersections and equal intersection spacing, features that favor natural coordination between intersections.

Lämmer and Helbing (2008) also propose decentralized, cycle-free control with rules that lead to self-organization. They develop an optimal switching rule considering a limited prediction horizon based on minimizing waiting time to queued vehicles and considering losses from switching. This leads to favoring movements with high discharge flows, that is, multi-lane movements compared to minor movements with a single lane. This requires introducing a “supervisory” set of rules to limit waits on minor movements. Vehicles expected to pass through on green without delay are not accounted for. Tests in limited, idealized cases show that coordination emerges.

2.1. Actuated control's self-organizing potential

Traditional actuated control is self-organizing at the intersection level in that it constantly adapts to the ideal cycle length for its input volumes, and because it naturally recovers from oversaturation and priority interruptions. In addition, actuated control logic naturally contains a weak coordination mechanism that arises from information carried by platoons released from an upstream signal. If the offset between two neighboring signals ever becomes such that a platoon discharged from the first signal arrives while the second signal is green, the second signal will hold its green until the platoon passes, and thus the two signals will end their green with an ideal progression offset. Suppose now that the cross-traffic demands at the two intersections are such that the following red interval at the first intersection is just as long as, or a little longer than, the second intersection's following red interval. Then the next platoon released from the first intersection will again arrive on green at the second intersection, maintaining synchronization.

However, this coordination mechanism will maintain synchronization only if signals have roughly the same natural cycle length. If the first intersection has a naturally shorter cycle than the second, then when offsets between the two intersections become close to ideal, actuated logic will lengthen the second intersection's already-longer cycle (holding until the platoon passes), without creating any tendency to lengthen the first intersection's cycle, making the intersections quickly diverge from synchronization. Or if the first intersection's natural cycle length is considerably longer than the second intersection's, then in the cycle after synchronization was achieved, the second intersection may cycle so quickly that the platoon released from the first intersection will arrive at the second intersection after its arterial phase has ended, instantly ending synchronization.

Gershenson (2005) and Gershenson and Rosenblueth (2012) have demonstrated that under ideal conditions, this weak coordination mechanism can bring about an amazing level of global coordination. Their test network is a grid of one-way streets, with alternating northbound and southbound streets crossed by alternating eastbound and westbound streets. The rules that govern each intersection's signal are similar to those of standard actuated control: they include an extension rule (hold the signal green while a platoon is passing), a gap-out rule (end the green if there is no more queued traffic and a competing traffic stream has a queue), and a spillback prevention rule (end the green if the downstream segment is blocked). They also add an interrupt rule (switch control if the accumulated waiting time in the waiting approach's queue reaches a certain threshold) which we have not adopted. Visualizations (Gershenson, 2014) show the strong coordination that is achieved, with platoons advancing with little delay through successive intersections in all four directions. They find that decentralized, actuated control results in less delay than traditional coordinated control, even though the test setting is ideally dimensioned for coordinated control, offering perfect green waves in all directions.

2.2. Research objective

Unfortunately, in real road networks, we do not see this self-organizing behavior emerge from actuated control. To some extent, Gershenson et. al.'s remarkable results are due to an idealized network in which the streets are all one-way, have equal average traffic demand on all streets in a given orientation (east–west or north–south), and have no turning traffic. Equal demands on all cross streets means that all intersections have the same natural cycle, which enables the previously described weak coordination mechanism inherent in actuated control. One-way streets make achieving coordination much easier than if streets are two-way; they also result in simple two-phase control at every intersection.

Additional coordination mechanisms are needed to deal with the realities of two-way streets, irregular intersection spacing, irregularly distributed traffic volumes, and multi-phase control. The goal of this research was to develop additional coordination mechanisms that could be applied within the framework of actuated control and to test them, with simulation, in a realistic traffic setting to see whether they would lead to self-organizing behavior and with it, improved performance.

3. Proposed control logic

This section describes the control logic, detector configuration, and communication protocols that we propose for self-organizing control. It is divided into four sections: logic for making standard actuated control efficient, rules for two

newly-proposed coordination mechanisms, secondary extension and dynamic coordination, and rules for oversaturated conditions.

3.1. Efficient actuated control

The control scheme we propose includes the standard elements of actuated control including a fixed service sequence, minimum and maximum green times, skipping minor phases that have no call, switching control when gaps are detected in the arrival flow, and switching control when spillback is detected on a departure leg.

At the same time, we specify settings and detection logic that maximize the efficiency of actuated control. Detector configuration and control settings can affect how quickly the controller detects when a traffic movement is no longer discharging at or near the saturation flow rate and switches control to the next phase. The lost time that occurs from not detecting and switching quickly (“snappy operation”) can substantially increase cycle length and delay (Furth et al., 2009). Four important elements of efficient control are using upstream detectors as extension detectors, short unit extensions, non-simultaneous gap-out, and short minimum green periods (Newell and Osuma, 1969; Furth et al., 2009). These features are all adopted in our tests. Upstream detectors are located about 2 s travel time before the stopline. We use a 6 s minimum green for turning movements; for through movements, the minimum green is 10 s when there is no pedestrian call. When there is a pedestrian call, minimum split (green plus yellow) is the pedestrian clearance time (crosswalk length divided by 3.5 ft/s) plus 7 s.

Cesme and Furth (2012) have shown that while traditional gap detection is effective on single lane approaches, it is not efficient on multilane approaches at detecting when saturation flow has ended. They showed that on multilane approaches, using multi-headways instead of simple gaps as the extension criterion allows one to discriminate more powerfully between saturation flow and below-saturation flow, allowing for a snappier operation with little risk of cutting off a discharging queue (e.g., on a three-lane approach, that might mean terminating the green unless three vehicles, regardless of lane, cross the stopline in the next 3.3 s.). The control logic we propose includes this feature.

Our proposed control method also counts the number of vehicles within the “trap” between the upstream and stopline detectors, incrementing the count with every upstream detection and decrementing with every stopline detection. Trap counts are a well-established feature of traffic control, used, for example, with so-called volume-density control (Koonce et al., 2008). They are used as a measure of queue length in several of the algorithms that follow.

3.2. Rules for secondary extension

The primary coordination mechanism we propose to add is the “secondary extension.” If a platoon is expected to arrive on an approach facing a green shortly after a detected gap, it may be worth extending the green – akin to green extension for transit – because such an extension can drastically reduce delay for the vehicles in the arriving platoon while increasing cycle length, and thereby delay, for other traffic by only a small amount.

Each intersection approach has an advance detector placed 20 s travel time upstream of the stopline or just after the immediately upstream intersection, whichever is closer. From that advance detector, using a given travel time offset between the detector and the stopline, the controller monitors the arrival profile at the stopline over a horizon of about 20 s. The arrival profile is the cumulative expected arrivals at a time t in the future for $t = 1, 2, \dots$ seconds (Fig. 1).

If the upstream intersection is less than 20 s distant, vehicles in the upstream intersection’s queue are added to the arrival profile when the upstream arterial phase is activated (a phase is “activated” at the onset of the change interval preceding its green), assuming a given saturation flow rate and start-up lost time. The proposed arrival prediction method ignores possible platoon dispersion.

Willingness to grant a secondary extension should increase as the set of arriving vehicles becomes larger, denser, or more imminent. A proposed measure that combines these three features is L_v^* , the lost time per vehicle, defined as the ratio of wasted green time during the secondary extension to the number of arrivals during the secondary extension, and minimized by considering different potential lengths of secondary extension. Let time t be initialized so that $t = 0$ at the first moment the gap-out criterion is satisfied, and let $L_v(t)$ = lost time per vehicle if the secondary extension’s length is t , given by,

$$L_v(t) = \frac{t - n(t) * h_{\text{sat}}}{n(t)} \quad (2)$$

where $n(t)$ is the number of vehicles expected to pass the stopline if the green phase is further extended by t and h_{sat} is the saturation headway. The numerator is the wasted green time, i.e., green time in excess of what would be needed to serve the arriving vehicles if they discharged at saturation headway. $L_v(t)$ is calculated for discrete values of t up to SX_{max} , the maximum allowed length of secondary extension, to find an optimizing value of t :

$$L_v^* = \min_{t \leq SX_{\text{max}}} L_v(t) \quad (3)$$

In the common case in which the arrival profile includes a gap followed by a dense platoon of uniform density, L_v^* will be minimized at the time t at which the last vehicle in the platoon reaches the stopline. Observe that in such a case, L_v^* will be smaller if the platoon is larger, denser, and more imminent (shorter time until it arrives).

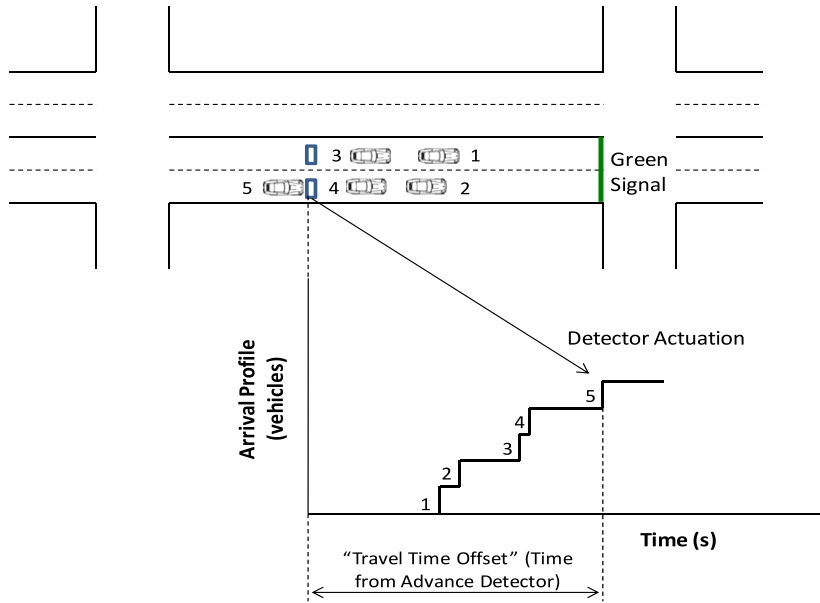


Fig. 1. Arrival profile for an approaching platoon for secondary extension logic.

Willingness to grant a secondary extension should also depend on how much excess capacity an intersection has – the more excess capacity, the more an intersection can afford to “waste” green time in order to improve progression. Affordable lost time $L_{\text{affordable}}$ (s/vehicle) is given by Eq. (4) and also illustrated in Fig. 2.

$$L_{\text{affordable}} = \min \left[2, 2 \left(\frac{1}{v/c} - 1 \right) \right] \tag{4}$$

The intersection’s volume to capacity ratio, v/c , shown in Eq. (4) can be calculated as follows:

$$v/c = \frac{\sum_{\text{critical}} \frac{v_i}{s_i}}{1 - L_{\text{sum}}/C} \tag{5}$$

where v_i and s_i are a movement’s arrival rate and saturation flow rate, the sum is over critical movements only, L_{sum} = sum of the lost time for the critical movements, and C = maximum desirable cycle length. Our tests used $C = 90$ s, assumed that L_{sum} was 4 s per critical phase, and measured v and s adaptively (i.e., using detectors within the simulation environment, updating estimates every 5 cycles).

As shown in Fig. 2, affordable lost time per vehicle is capped at 2 s for $v/c \leq 0.5$, and then declines with v/c , making it more difficult for an arriving platoon to secure a secondary extension, becoming 0 when $v/c = 1$. The decline is linear with $1/(v/c)$. When an intersection becomes over capacity, secondary extensions are inhibited entirely. Sensitivity analysis (Cesme, 2013)

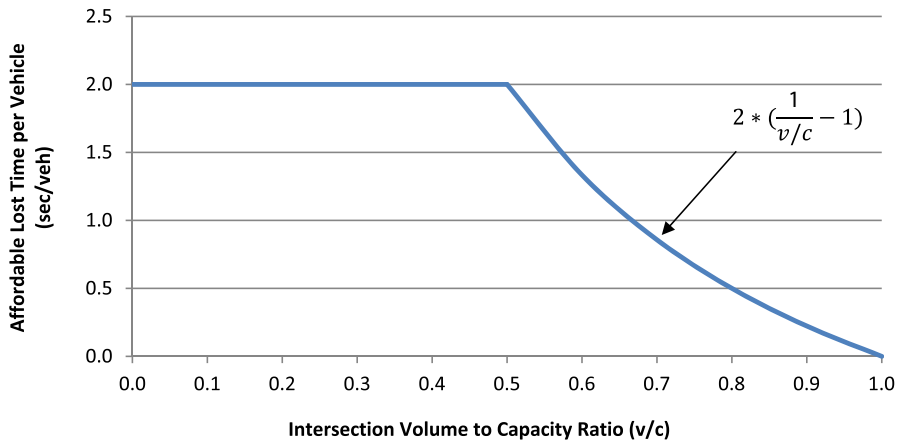


Fig. 2. Affordable lost time per vehicle versus intersection volume to capacity (v/c) ratio.

with respect to the parameters used in the affordable lost time function as well as the shape of the declining part of the function (linear, convex, and concave with respect to v/c) found little difference in average delay.

The maximum scheduled secondary extension, SX_{\max} , is 20 s for the phase serving the critical arterial through movement. This generous limit was given because lengthening a critical phase to serve a temporary surge of demand in one cycle is likely to lead to a shorter next cycle, thus compensating traffic streams that will experience a longer red due to the secondary extension. Sensitivity analysis to SX_{\max} was performed, in which self-organizing signals were also tested with 10 s or 30 s of maximum secondary extension (Cesme, 2013). Results showed applying 10 s of maximum secondary extension resulted in higher average delays (2% increase in average delay compared to 20 s of maximum secondary extension). Applying 30 s of maximum secondary extension reduced delay marginally (less than 0.5% compared to 20 s). However, it is important to note that achieving a 30 s prediction horizon may not be practical, particularly along arterials with closely spaced intersections. Therefore, maximum secondary extension was set to 20 s.

Because compensation cannot thus be expected when making extensions for non-critical movements, a more stringent maximum is specified for the arterial through movement in the non-critical direction:

$$SX_{\max} = \min\{\max(10, \Delta C_{n1}, \Delta C_{n2}), 20\} \quad (6)$$

where ΔC_{n1} and ΔC_{n2} are the difference between its neighboring intersections' needed cycle length and its own. Eq. (6) limits the secondary extension for the non-critical arterial phase to 10 s when the subject intersection's needed cycle length is close to that of its neighboring signals. But if a neighboring signal has a substantially longer needed cycle length than the subject intersection, then a longer extension will be allowed, because a longer extension should improve coordination by bringing cycle lengths closer to equal. At the moment an arterial phase gap is detected, $L_v(t)$ is calculated for candidate extension lengths $t = 2, 4, \dots, SX_{\max}$; the minimum with respect to t , L_v^* is identified along with the minimizing argument, t^* . If it meets the affordable lost time criteria for the intersection's current v/c , a secondary extension will be granted, with scheduled length t^* ; otherwise the phase gaps out.

Once a secondary extension reaches a duration t^* , normal gap-out logic governs; that is, the green will extend beyond t^* until low flow is detected, subject to a maximum green. That way, an entire platoon can be served even if all of it was not detected because of a limited horizon. A phase may receive only one secondary extension per cycle. As a rule, generous maximum greens were used in order to avoid over-constraining the system.

3.3. Dynamic coordination for coupled intersections

Where intersections are closely spaced, a critical weakness of local actuated control is that a failure to provide good progression during high demand periods can cause spillback or starvation, dramatically reducing intersection capacity. Even when there is no threat of overcapacity, short stretches with closely spaced intersections offer an opportunity for good two-way progression by having arterial greens begin simultaneously. On short stretches, progression can also benefit transit when there is no intermediate stop.

Therefore, our proposed control logic groups small sets of closely spaced intersections into "coupled zones" in which signals communicate with one another so that they cycle synchronously, without imposing a fixed cycle length. This approach, called "dynamic coordination," was inspired by Zurich's traffic signal control, which divides its network into zones of one to three closely-spaced intersections separated by segments long enough that queuing there will not cause spillback (Furth et al., 2010). Using dynamic cycle lengths keeps cycles as short as the traffic demand will allow, and provides flexibility for transit priority, with green waves for vehicles simply following buses and trams. The proposed logic is not suitable for tight downtown grids with long stretches of closely spaced intersections.

We followed the rule of thumb that intersections were coupled when less than 500 ft and not coupled when spaced more than 600 ft apart. Within the range of 500–600 ft, discretion was used depending on the number of intersections involved in the coupled zones and the likely cycle lengths (which affect queue lengths).

Each coupled zone has a critical intersection, the intersection with the greatest needed cycle length, updated adaptively based on volume and saturation flow measurements over the last five cycles. A zone's "governing phase" is the critical arterial phase at the critical intersection. The coordination objective when the v/c ratio at the critical intersection is below 0.9 is to have good two-way progression by having arterial through phases in both directions begin (roughly) simultaneously; above 0.9, it's to provide ideal progression for the critical direction, protecting the governing phase from both starvation and spillback.

Where the arterial has left-turn phases, two-way progression can be improved by sequencing phases with leading lefts entering a coupled zone and lagging lefts leaving the zone (implying leading through movements entering and lagging through movements leaving the zone). This sequencing reduces the chance of starvation at the critical intersection. Lagging lefts at the end of a zone also provide better progression for vehicles turning left at the end of the zone. Phase sequencing is taken as an external input.

At the critical intersection, a "queue-adjusted earliest activation time" is calculated for the governing phase at every phase transition. It represents the earliest time the standing queue for that phase could clear, minus the preceding change interval. The earliest activation time of the governing phase can be calculated based on the current signal state, commitments made by the local controller such as minimum green or pedestrian clearance, and the minimum duration of intervening phases, which is calculated based on the length of the standing queue using a given lost time and saturation headway.

Intervening phases that do not have a call are expected to be skipped. “Queue-adjusted earliest activation time” is the earliest activation time of the governing phase plus the expected time to clear its standing queue, the latter being the product of its standing queue length and a given saturation headway. The critical intersection communicates its queue-adjusted earliest activation time to other intersections in its zone.

When the critical intersection v/c is less than 0.9, this queue-adjusted earliest activation time becomes the scheduled activation time for arterial through phases at the non-critical intersections (Scheduled activation time can be understood as the scheduled green end time for the phase immediately preceding an arterial phase.). Non-critical intersections will usually be ready to activate their arterial through phases before the scheduled activation time, because non-critical intersections usually need less time to cycle than the critical intersection. Where the arterial through phase is immediately preceded by a cross-street through phase, a “green hold” (not allowing the phase to yield its green) is put on the cross-street phase until its scheduled green end so that arterial through phases in the coupled zone start their green at the same time (Fig. 3a). In this way, the slack time forced by dynamic coordination is used to reduce delay on the cross-street and reduce cross-street demand for the next cycle, thereby creating additional flexibility for the following cycle.

Where an arterial through phase at a non-critical intersection is immediately preceded by an arterial leading left phase, giving all the slack time to that leading left is not as productive. It is preferred to give the slack time to the cross street through phase (Fig. 3b). A scheduled green end time for the cross street through phase is then determined by estimating the needed split for the intervening left turn based on its queue count, amplified by 10% to allow for late arrivals.

When the critical intersection v/c exceeds 0.9, cross street green is truncated at its scheduled end time. When critical intersection v/c is less than 0.9, the cross street green will terminate at its scheduled end green time unless one or both cross street directions have not yet registered a gap. In such a case, the cross street phase is not truncated; it is allowed to terminate normally (i.e., by gap-out or max-out). In this way, we preserve the feature of actuated control that fluctuations in demand will not result in overflow except in the case of max-out.

When the critical intersection's v/c exceeds 0.9, the scheduled activation time for the critical direction arterial phase at non-critical intersections is the queue-adjusted earliest activation time of the governing phase plus or minus the travel time to/from the critical intersection (plus if it's downstream of the critical intersection, minus if upstream). These offsets are meant to give the critical direction perfect progression through the zone, without any spillback or starvation affecting the governing phase. Note that at non-critical intersections, arterial phases are biased to begin early, since they use the “earliest” activation time of the governing phase. If arterial phases upstream of the critical intersection begin early, there may be a short period of spillback at those upstream intersections, but there will be no starvation at the critical intersection. And if arterial phases downstream of the critical intersection begin early, there may be temporary starvation at those downstream intersections, but there will not be spillback blocking the critical intersection.

As stated earlier, when the critical intersection's v/c exceeds 0.9, cross-street through phases at non-critical intersections are truncated at their scheduled green end time. Priority is given to avoiding spillback or starvation that could reduce throughput for the governing phase, because overcapacity at the governing phase has a lasting effect throughout the zone. Non-critical intersections have slack capacity that will allow any cross-street overflow to clear in later cycles.

There is a chance that an arterial phase at a non-critical intersection could begin so early that it gaps out before the arrival of the platoon released from an upstream intersection within the zone. To prevent this kind of early termination, a “green hold” is applied to an arterial phase at a non-critical intersection from its scheduled activation time until the expected arrival time of its upstream platoon (start of arterial green at the upstream intersection plus a given travel time). Once the platoon arrives, normal gap-out logic will keep the phase green until the platoon has passed.

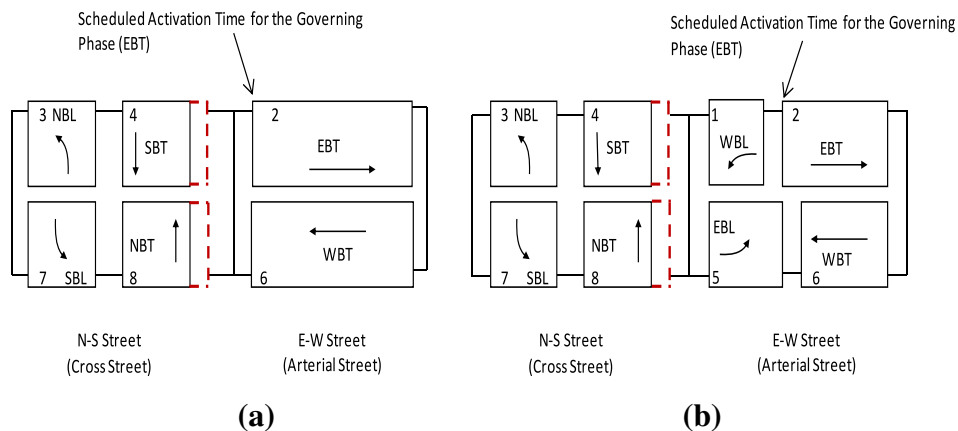


Fig. 3. Green hold on the cross street through phase at a non-critical intersection to provide simultaneous green for arterial through phases – (a) Arterial through phase is preceded by a cross-street through phase; (b) arterial through phase is preceded by an arterial leading left turn phase.

This dynamic coordination logic will still allow double cycling in the case that an arterial phase at a non-critical intersection begins its green so early that it gaps out before its scheduled activation time. Simulation experiments described later in this paper show occasional double cycling within coupled zones.

3.4. Self-organizing control during temporary oversaturation

During temporary periods of oversaturation as well as the periods leading up to them and recovering from them, the main objective of traffic signal control should shift to maximizing throughput. By continually monitoring input and discharge volumes, the v/c ratio of each intersection is updated every cycle, using a 5–previous cycle rolling horizon. As explained earlier, the secondary extension rule, which “wastes” some capacity in order to improve progression, becomes very stingy as intersection v/c approaches 1.0, and is inhibited during periods of oversaturation. Standard actuation logic, which switches control when discharge falls significantly below the saturation flow rate, will then keep discharge near its maximum, leading to the desired overall behavior.

Queue interactions can sometimes cause upstream blockage (starvation) or downstream blockage that limit throughput. The desired overall behavior is to (1) respond to blockage by switching control to a phase that is not blocked and (2) limit the formation of blockage. The usual gap-out logic will detect and respond to starvation unless constrained by minimum green. Self-organizing control also includes spillback detectors, located just downstream of the intersection, that trigger the end of green (subject to minimum green) when downstream blockage is detected.

We also include logic for dealing with blockage involving pocket turning lanes of limited length. If the queue in a pocket lane spills back, it will block a through lane, decreasing discharge flow for the through movement. This is most critical for a left turn pocket on the approach with the arterial critical direction, and has only been applied in this direction (application to the other direction is a topic for future research). A pocket spillback detector is located at the rear of the left turn pocket. If the left turn is a lagging left, detection of spillback will truncate the opposing through movement, subject to minimum green, and begin the lagging left phase early, in order to minimize blockage to the neighboring through lane. If the left turn is a leading left, pocket spillback will trigger a reservice phase, meaning the left turn movement will get a lagging interval as well as a leading interval, again by truncating the opposing through movement. In a simulation experiment reported in [Cesme and Furth \(2013\)](#), the reservice phase was called in 70% of the cycles, and its implementation increased critical intersection throughput by 183 vehicles per hour per lane, greatly reducing the length and severity of a temporary period of oversaturation.

Because the treated left turn and its opposing through movement are not critical, they have excess capacity that can, to some extent, absorb green truncations and the additional lost time involved in reservice. If repeated truncation of the opposing through movement results in long queues, that movement will become recognized as the critical movement, and will then be protected from truncation.

For situations in which intersection spacing is so short that queues at one intersection are apt to affect a neighboring intersection, some deliberate coordination is needed to ensure maximum throughput at the critical intersection. We have already described the proposed dynamic coordination logic for this situation. When v/c at the critical intersection is below 0.9, some capacity is “wasted” in order to provide good two-way progression by scheduling simultaneous green start times for arterial phases in both directions. When v/c exceeds 0.9, offsets are calculated to provide progression in the critical direction that avoids starvation and spillback at the critical intersection by allowing minor starvation at downstream intersections and minor spillback at upstream intersections.

Our proposed method of self-organizing control in relation to oversaturation is described at greater length in [Cesme and Furth \(2013\)](#), which also describes several simulation tests. In one benchmark test, self-organizing control led to 45% less delay than coordinated control. In realistic network simulations, delay reductions of 8% and 35% were observed.

4. Simulation tests

The proposed control logic was tested in a microsimulation environment using VISSIM. Control logic, programmed in C++, was run in parallel using VISSIM's application programming interface in which at every time step, detector information is passed from the simulation to the control program, which in turn passes information about signal states back to the simulation program.

The proposed self-organizing logic was compared with coordinated–actuated control on two arterials that currently operate under coordinated–actuated control. One is Beacon Street in Brookline, Massachusetts, with irregular intersection spacing, light rail transit (LRT) operating in a median reservation, and heavy pedestrian cross traffic. The second is Rural Road in Tempe, Arizona, with little traffic other than private motor vehicles and well-spaced intersections that make it amenable to traditional coordinated control.

4.1. Test setting 1: Beacon Street, Brookline, Massachusetts

Beacon Street ([Fig. 4](#)) is a divided, 4-lane arterial with a 30 mph speed limit and LRT operating every 7 min in a median reservation. The modeled section is 1.3 miles long and has 12 traffic signals with spacing ranging from 280 to 970 feet, two of which are for midblock pedestrian crossings. LRT is not controlled at the two midblock signals, and so LRT results are

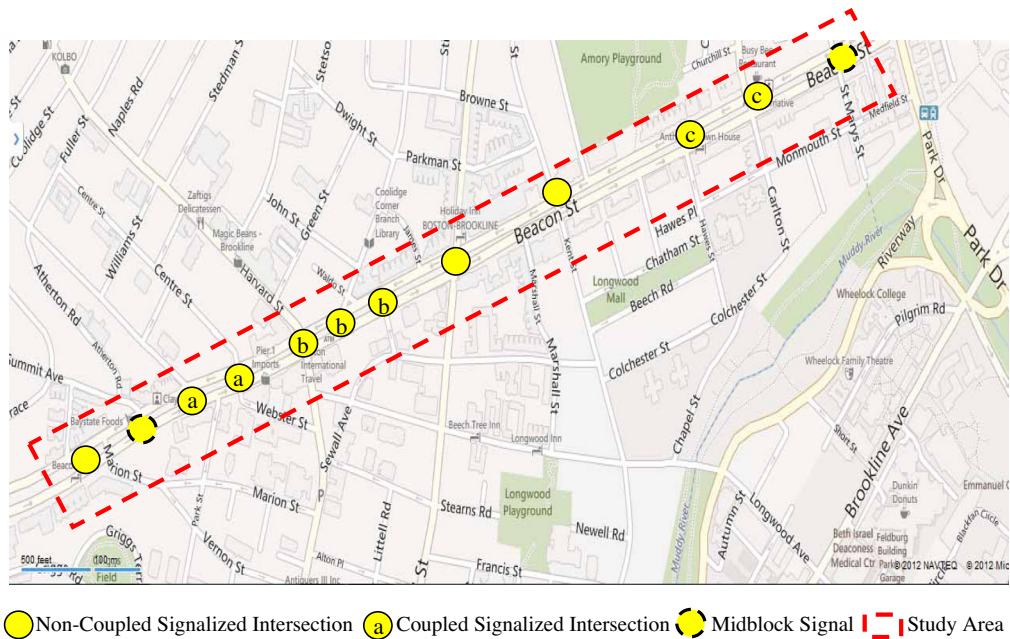


Fig. 4. Beacon Street corridor, Boston, Massachusetts.

presented as results over 10 intersections. The period modeled is the a.m. peak, for which the peak direction is eastbound, also called inbound. Traffic counts and signal timing parameters came from a 2009 study by Vanasse Hangen Brustlin, Inc. Transit schedule and passenger on-off data were obtained from the transit agency and field studies.

Pedestrian signals are actuated, but demand for crossing Beacon Street exceeds 100 pedestrians per hour at all crossings except Hawes Street, so pedestrian phases are served almost every cycle. All transit stops were modeled as far-side in order to make them more amenable to transit priority, even though the existing layout has a mix of far-side and near-side stops. To model the random arrival time and load of trains at the first stop of the study segment (which is not the beginning of the line in either direction), dummy stops were placed at the far ends of the segment with random dwell times. To model the additional boarding time caused by vehicle crowding, transit vehicle loads were tracked in VISSIM, and dummy traffic signals that apply only to trains were placed just after each stop to hold the train for a computed crowding penalty calculated as in Wadjas and Furth (2003).

For self-organizing control, the corridor was configured with three coupled zones. From the west, the zone a covers two signals (Webster, Winchester), zone b has three signals (Harvard, Pleasant, Charles), and zone c has two signals (Hawes, Carlton). The remaining five full signals as well as two midblock signals (for pedestrian crossings) are uncoupled.

The existing coordinated-actuated control, which had been optimized for traffic projections expecting strong traffic growth, operated at the time of the study with a 90-s cycle. Because the 2009 counts show mainline volumes nearly 20% below those projections, timing parameters were re-optimized using Synchro, allowing both leading and lagging left turns, leading to a common cycle length of 80 s. Manual adjustments were made to some offsets in order to improve progression based on observing the simulation.

With an 80 s cycle length, the critical intersection operates at 81% of capacity. In order to model performance under varying demands, volumes were scaled so that the critical intersection v/c was 0.67 during the warm-up (time 0–15 min) and first analysis period (time 15–30), 0.81 during the second period (time 30–45), 0.95 during the third period (time 45–75), and 0.67 during the fourth period (time 75–105). Five simulation runs were made for each control scenario, with results reporting averages for the 90-min period following warm-up.

Table 1 compares impacts to general traffic and to transit. All comparisons that follow use as a base case coordinated-actuated control without transit priority, optimized for period 2 volumes. Without transit priority, self-organizing control reduces overall vehicular delay by 14%. Delay for vehicles traveling along the arterial increases by about 11%, coupled with decrease in cross-street delay and turning traffic delay, which in turn results in lower overall vehicular delay. During time period 2, the period with existing traffic volumes, average cycle length falls from 80 to 69.2 s. Pedestrian delay in the network for all time periods also decreases by 12% as a result of shorter cycle lengths. Average transit delay increases slightly, by 1.0 s per intersection (both directions combined).

Signal priority under both self-organizing and coordinated-actuated control included green extension and red truncation. Check-in detectors were located approximately 650 ft upstream of stop line, corresponding to 15 s of travel time, or just downstream of the previous intersection where intersection spacing is less than 650 ft. Maximum green extension was then

Table 1

Simulation results for Beacon Street with and without transit signal priority.

	Coordinated–actuated ^a		Self-organizing		
	No TSP	TSP	No TSP	TSP	Conditional TSP ^b
Average network delay (s/vehicle) and change from base ^c	68.4 ^c (0%)	74.0 (8%)	58.6 (–14%)	67.1 (–2%)	70.5 (3%)
Average pedestrian delay (s/pedestrian) and change from base	55.5 (0%)	55.5 (0%)	48.8 (–12%)	48.2 (–13%)	Not measured
Average number of stops and change from base	1.74 (0%)	1.80 (3%)	1.80 (3%)	1.94 (11%)	Not measured
Average arterial delay (s/vehicle) and change from base	111.7	107.5 (–4%)	124.3 (11%)	134.4 (20%)	Not measured
Inbound (peak direction) train delay (s/train/intersection) and change from base	19.8 (–)	14.9 (–4.9 s)	21.7 (1.9 s)	6.4 (–13.4 s)	9.9 (–9.9 s)
Overall train delay (s/train/intersection) and change from base	20.2 ^c (–)	13.7 (–6.5 s)	21.2 (1.0 s)	7.1 (–13.1 s)	9.9 (–10.3 s)
Percent of trains eligible for priority (only late trains request priority)	0%	100%	0%	100%	69%
Inbound train headway coefficient of variation at first stop in modeled network	0.426	0.426	0.426	0.426	0.426
Inbound train headway coefficient of variation at last stop in modeled network	0.580	Not measured	0.562	Not measured	0.437
Average cycle length (s) during base period ($v/c = 0.81$)	80.0	80.0	69.2	Not measured	Not measured

^a With cycle length, splits, and offsets optimized.^b Also includes cross street interruptions at Harvard Street.^c Base case is coordinated–actuated without TSP.

15 s or travel time from the upstream intersection, whichever is smaller. Note that red truncation has limited applicability in this context because it cannot interrupt pedestrian clearance. Under self-organizing logic, two additional priority tactics, phase rotation and phase insertion, were also applied. Phase rotation means switching arterial lefts from leading to lagging and vice versa, and phase insertion means inserting a short transit-only phase after a lagging left, as described in Furth et al. (2010). General traffic is not released during such an inserted transit-only phase, even if it runs parallel to transit, in order to discourage pedestrians from beginning to cross.

With priority applied to all trains, TSP reduced transit delay by 6.5 s per intersection when applied within a framework of coordinated–actuated control, and by 13.0 s when applied within a framework of self-organizing control. The improved performance with self-organizing control is mainly attributable to the more aggressive priority tactics that could be applied in a framework free from a fixed cycle length. With coordinated–actuated control, TSP increases overall vehicular delay by 8.2% compared to the base case. With self-organizing logic, TSP also increases overall vehicular delay, but it remains 2% lower than overall vehicular delay in the base case.

Self-organizing control was also tested in a “conditional priority” scenario with a goal of limiting bunching in the inbound (peak) direction by withholding priority from trains that were too close to their leader. Inbound, priority was given to trains only if their headway exceeded 315 s, or 75% of the scheduled headway. Outbound, the threshold was only 180 s, because in the segment modeled the outbound direction is dominated by alightings, and alighting passengers are not hurt when trains run early. By giving priority to late trains but not to early trains, conditional priority helps limit crowding and improve service reliability while at the same time creating less disruption for other traffic (Furth and Muller, 2000). In order to further test the system’s ability to heal from priority interruptions, the conditional priority scenario also includes granting a 20 s green extension for cross-street traffic at Harvard Street once every 6 min, as might occur from giving priority to a crossing bus route. In case of conflicting priority requests between LRT and the cross-street bus, priority was given to LRT.

In the conditional priority scenario, 69% of trains request priority. Average transit delay falls by 10.3 per intersection, or 14.9 s if averaged only over trains requesting priority. The effect of conditional priority on bunching is indicated by the headway coefficient of variation (cv) for inbound service. Bunching is a self-reinforcing phenomenon as late trains, facing more waiting passengers than usual and with a boarding process slowed by crowding on the vehicle, become ever later. Without priority, headway cv grows from 0.43 at the start of the study segment to roughly 0.57 at the last stop under both self-organizing and coordinated–actuated control. With conditional priority, headway cv remains almost constant along the segment. Overall vehicle delay shows a small increase, probably due to the cross-street priority interruptions that are not part of the base case.

For more insight into the behavior that emerges from self-organizing logic, Fig. 5 shows a box-plot of cycle lengths (minimum, 25-percentile, 75-percentile, and maximum) by intersection during periods 3 and 4, without signal priority. Distinct borders shown in Fig. 5 indicate coupled zones. For a given intersection and period, cycle length varies widely, showing the responsiveness of the logic to traffic. Cycle length tends to be greater in periods of greater traffic volume. Cycle length also varies systematically from intersection to intersection, but that variability is limited because of how intersections synchronize with one another. Variability between intersections is the greatest during the period of greatest demand, with the critical intersection, St. Paul, cycling every 94 s on average compared to 78 s for the other intersections. The long average cycle length at Kent, whose own traffic doesn’t demand a long cycle, shows that secondary extension logic leads it to synchronize with its neighboring intersection, St. Paul, in many – but not all – cycles.

Within each coupled zone, cycle lengths at the member intersections have almost the same distribution, as expected. The only exception is the coupled zone Hawes–Carlton during period 3, for which Hawes cycles three times more per hour than Carlton, evidence of occasional double cycling.

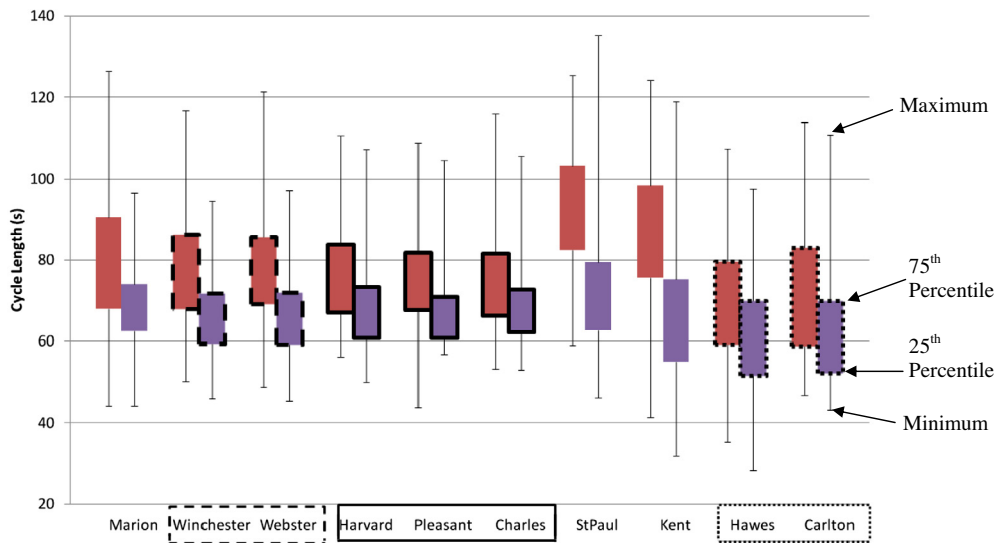


Fig. 5. Distribution of cycle length by intersection for time periods 3 and 4. Note: Plot shows minimum, 25-percentile, 75-percentile, and maximum cycle length for each intersection during time period 3 (left, in red) and time period 4 (right, in violet). Coupled zones are indicated by distinct borders. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Test setting 2: Rural Road, Tempe, Arizona

A 3.1 mile section of Rural Road in Tempe, Arizona between Warner Road and Minton Drive was the second test site (Fig. 6). It is better suited than Beacon Street to traditional two-way coordination, with intersections regularly and distantly spaced, few pedestrians, and little transit service. It has two to three lanes per direction as well as a median left-turn lane, and a speed limit of 45 mph. Traffic counts and signal timing information were obtained from Maricopa Association of Governments (MAG).

Morning peak traffic was modeled with two adjustments. First, because we wanted to test algorithmic performance in undersaturated conditions, traffic volumes were scaled down to make the v/c ratio at the critical intersection equal 0.9 (Tests with actual morning peak traffic volumes, for which performance is heavily influenced by a control logic's ability to handle oversaturation, are described in Cesme and Furth (2013) and in Cesme (2013)). Second, in order to test performance with frequent transit priority interruptions, we modeled the bus headway along the arterial as 10 min instead of 20 min, and modeled all stops as far-side. Pedestrian phases were actuated. Because pedestrian volumes were unknown, we assumed pedestrian demand of 30 peds/h at major intersections and 15 peds/h at minor intersections.

In the transit priority scenario, priority tactics included green extension with a maximum extension of 15 s and early green, in which left turn phases can be truncated after six seconds and through phases after ten seconds, or, if there was a pedestrian call, after the ped minimum green. With self-organizing control, an additional priority tactic was phase rotation, in which a leading left is changed to a lagging left.

For coordinated-actuated control, timing parameters were optimized using Synchro, yielding a 100 s cycle length (50 s at Carver). For self-organizing control, only one coupled zone was specified, containing the two northernmost intersections, which are 595 ft apart. Simulation results are reported for a 60 min period of uniform demand following a 15 min warm up period, with averages reported from five simulation runs.

Overall performance is shown in Table 2. Compared to coordinated-actuated control, self-organizing control lowers average cycle length from 93 to 65 s. Not surprisingly, coordinated-actuated control provides better service for vehicles traveling along the arterial, resulting in fewer stops and in 17 s less delay, or about 2 s per intersection, for those vehicles that travel the full length of the corridor. However, when all vehicles are accounted for, self-organizing control results in 9% less vehicular delay, because its shorter cycle lengths and responsive green times serve turning and cross-street traffic so much better. It also results in 3% less transit delay and 30% less pedestrian delay.

Compared to coordinated-actuated control without priority, priority applied within a framework of coordinated-actuated control reduces transit delay by 9.6 s per intersection, while priority applied within a framework of self-organizing control yields a reduction of 12.3 s per intersection. The self-organizing framework makes priority especially more effective in the southbound direction, the direction less favored by arterial coordination. With either control paradigm, transit priority increases network delay by only 1–2 s per vehicle.

To get a better idea of algorithmic performance without the complication of pedestrians and transit, another scenario was modeled in which pedestrian demands were ignored and there was no transit priority. In this scenario, the optimal cycle length for coordinated-actuated control was 80 s except at Carver, which double cycles with a 40 s cycle length. Table 3

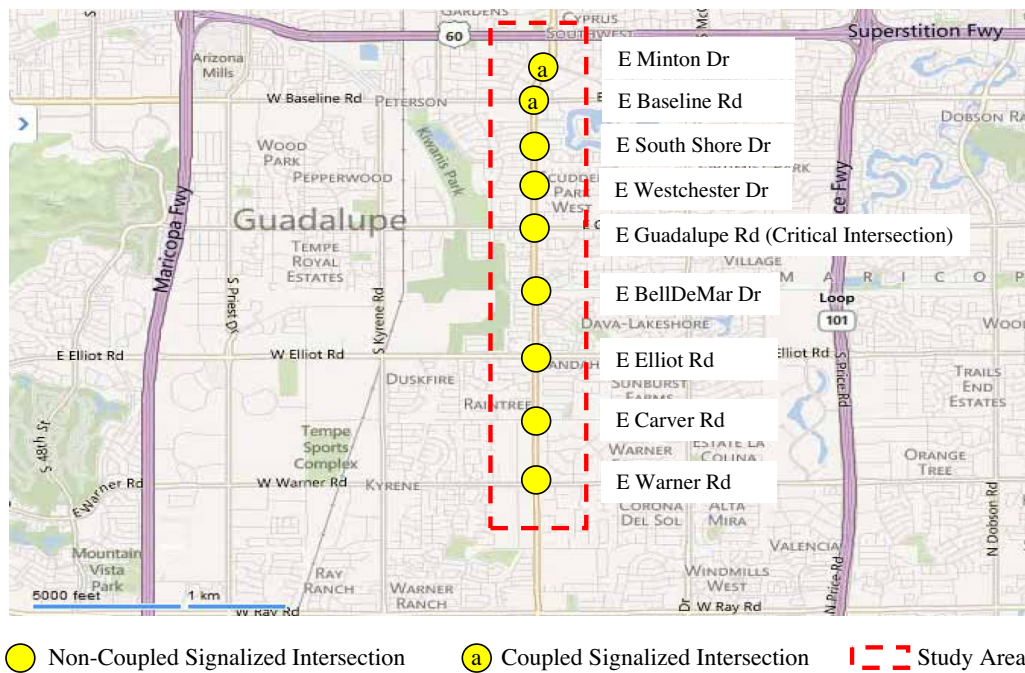


Fig. 6. Rural Road corridor, Tempe, Arizona.

Table 2
Simulation results for Rural Road with and without transit signal priority.

	Optimized coordinated–actuated		Self-organizing	
	No TSP	TSP	No TSP	TSP
Average cycle length (s)	92.9	92.9 (0% ^a)	65.1 (–30% ^a)	64.7 (–30% ^a)
Average network delay (s/vehicle)	54.4	56.7 (5%)	49.4 (–9%)	50.2 (–8%)
Average pedestrian delay (s)	45.1	44.7 (–1%)	31.5 (–30%)	33.6 (–26%)
Average number of stops	1.25	1.28 (3%)	1.33 (7%)	1.33 (7%)
Average arterial delay (s/vehicle)	114.3	113.8 (–0%)	131.4 (15%)	126.6 (11%)
Bus delay, northbound (s/bus)	198.4	100.7 (–49%)	202.7 (2%)	92.4 (–53%)
Bus delay, southbound (s/bus)	188.2	114.4 (–39%)	170.2 (–10%)	73.2 (–61%)
Percent of buses eligible for priority	0%	100%	0%	100%
Change in overall bus delay (s/bus/intersection)	–	–9.6 (–45%)	–0.8 (–3%)	–12.3 (–57%)

^a Values in parentheses indicate percentage change compared to optimized coordinated–actuated control without TSP.

shows average cycle length by intersection under self-organizing control. Averaged over all intersections, average cycle length falls by 20%, from 75.6 to 60.0 s. Average cycle length is almost unchanged at Guadalupe, the critical intersection, at the Baseline–Minton coupled zone, and at Carver, the intersection that double cycles under coordinated control. Other intersections, in contrast, cycle much faster than allowed with coordinated control, with average cycle lengths ranging from 47 to 63 s instead of 80 s. Average network delay with self-organizing control is 7.1% less than with the optimized actuated-coordinated plan (46.0 s versus 49.5 s).

For a more detailed look into how self-organizing control performs under this scenario, Fig. 7 shows a record of signal state changes (red and green intervals) on a segment with four intersections – major intersections at the two ends of the segment whose average cycle lengths are 80.0 and 78.6 s and where a majority of the cycle is red, and two minor intersections in the middle which cycle on average every 48.0 and 51.0 s, respectively, and where most of the cycle is green. In the coordinated plan, all four intersections run with an 80 s cycle. With self-organizing control, both green interval length and cycle length vary widely from cycle to cycle due to snappy operation parameters (which allow short greens) and secondary extension logic (which enables long greens). Diagonal lines show the leading edge of green bands at the progression speed. Most of them progress through at least two intersections; about half progress through three or more. Within this small sample, the number of through bands passing through all four intersections is 6 northbound (the peak direction in the a.m.) versus only 1 southbound, indicating that self-organizing controls “knows” that it’s more important to provide good progression in the peak direction.

Table 3
Average cycle length at the study intersections along Rural Road under self-organizing logic.

Intersection	Cycle length(s)	Change in cycle length (s)
E Minton Dr	78.3	-1.7
E Baseline Rd	78.6	-1.4
E South Shore Dr	51.0	-29.0
E Westchester Dr	48.0	-32.0
E Guadalupe Rd	80.0	-0.0
E Bell De Mar Dr	47.0	-33.0
E Elliot Rd	60.4	-19.6
E Carver Rd	39.9	-0.1
E Warner Rd	62.6	-17.4
Average	60.6	-15.0

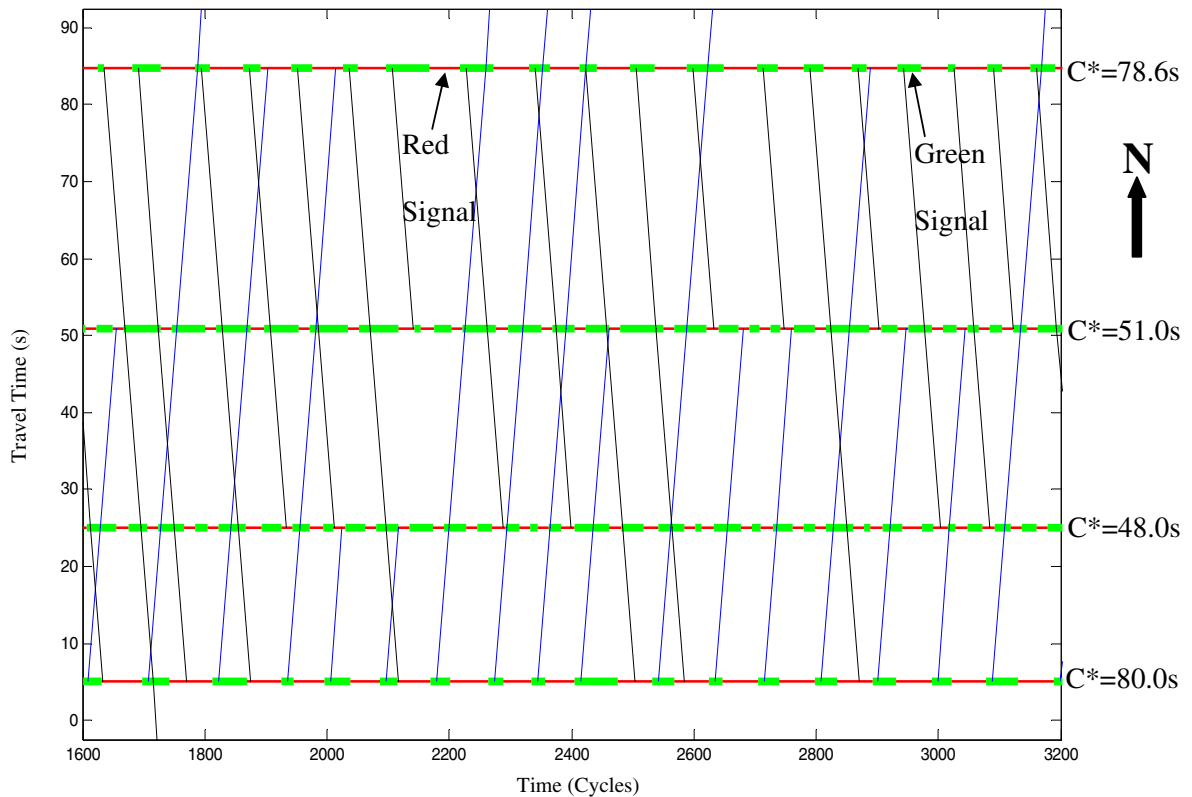


Fig. 7. Signal state changes and green bands for the critical (northbound, shown by diagonal solid lines) and non-critical (southbound, diagonal dashed lines) arterial direction for four intersections on rural road under self-organizing control logic. Note: C^* indicates average cycle length.

It is interesting to see that the minor intersections display exactly the kind of behavior hoped for. Most of the time, they double cycle with respect to the major intersections; however, sometimes they hold for a neighboring major intersection and match its longer cycle length.

5. Conclusions

Actuated traffic signal control logic has many advantages because of its responsiveness to traffic demands, short cycles, effective use of capacity, and amenability to aggressive transit priority. It also has some self-organizing properties that are very effective under ideal condition. However, under realistic conditions, it fails to provide good progression along arterials. At the same time, the traditional way of providing progression along arterials, coordinated control with a common, fixed cycle length, has many drawbacks stemming from its long cycle lengths, inflexibility in recovering from priority interruptions, and ineffective use of capacity when facing oversaturation.

Rules for secondary extension and dynamic coordination of short zones of closely spaced intersections have been proposed that add coordination mechanisms to traditional actuated control. They allow decentralized, actuated signal control

achieve a degree of synchronization that improves progression on arterials while preserving most of the benefits of actuated control. Tests show that the proposed self-organizing logic reduces overall delay to vehicles, enables transit priority to be far more effective without overly disrupting other traffic, and significantly reduces pedestrian delay, even on arterials with spacing that is well suited to coordinated–actuated control. It is also shown to be far more effective than coordinated–actuated control at minimizing the negative effects of temporary oversaturation.

American practice with arterial signal control, which almost exclusively relies on coordinated or coordinated–actuated logic, has long been criticized for focusing on the needs of arterial through traffic to the detriment of pedestrians, transit, crossing traffic, and traffic safety. Self-organizing signal control logic offers a promising alternative that better balances the needs of all users, reducing delay to pedestrians and allowing transit to be more effectively prioritized, while still providing reduced overall vehicular delay.

Acknowledgments

This research was supported by the National Science Foundation under Grant No. CMMI 1130936. The authors also would like to thank Maricopa Association of Governments, the Town of Brookline, and Caliper Corporation for providing data.

References

- Abu-Lebdeh, G., Benekohal, R.F., 2000. Signal coordination and arterial capacity in oversaturated conditions. *Transp. Res. Rec.* 1727, 68–76.
- Camazine, S., Deneubourg, J.L., Franks, N.R., Sneyd, J., Theraulaz, G., Bonabeau, E., 2001. *Self-Organization in Biological Systems*. Princeton University Press.
- Campbell, R., Skabardonis, A., 2014. Issues that reduce the performance of adaptive control systems during oversaturated conditions. In: *Transportation Research Board Annual Meeting 14-5438*, and to appear in *Transportation Research Record*.
- Cesme, B., 2013. *Self-Organizing Traffic Signals for Arterial Control*. Civil Engineering Dissertations. Paper 20. <<http://hdl.handle.net/2047/d20003105>>.
- Cesme, B., Furth, P.G., 2012. Multiheadway gap-out logic for actuated control on multilane approaches. *Transp. Res. Rec.* 2311, 117–123.
- Cesme, B., Furth, P.G., 2013. Self-organizing control logic for oversaturated arterials. *Transp. Res. Rec.* 2356, 92–99.
- Denney, R.W., Curtis, E., Head, K.L., 2009. Long green times and cycles at congested traffic signals. *Transp. Res. Rec.* 2128, 1–10.
- Dion, F., Rakha, H., Zhang, Y., 2004. Evaluation of potential transit signal priority benefits along a fixed-time signalized arterial. *J. Transp. Eng.* 130 (3), 294–303.
- Federal Highway Administration, 2009. *Manual on Uniform Traffic Control Devices*. Washington, DC.
- Furth, P.G., Muller, Th.H.J., 2000. Conditional bus priority at signalized intersections: better service quality with less traffic disruption. *Transp. Res. Rec.*, 173, 23–30.
- Furth, P.G., Cesme, B., Muller, Th.H.J., 2009. Lost time and cycle length for an actuated traffic signal. *Transp. Res. Rec.* 2128, 152–160.
- Furth, P.G., Cesme, B., Rima, T., 2010. Signal priority near a major bus terminal: case study of Boston's Ruggles Station. *Transp. Res. Rec.* 2192, 89–96.
- Gershenson, C., 2005. Self-organizing traffic lights. *Complex Systems* 16 (1), 29–53.
- Gershenson, C., Rosenblueth, D.A., 2012. Adaptive self-organization vs. static optimization: a qualitative comparison in traffic light coordination. *Kybernetes* 41 (3/4), 386–403.
- Gershenson, C., 2014. <<http://turing.iimas.unam.mx/~cgg/sos/SOTL/SOTL.html>> (accessed February 9 2014).
- He, Q., Head, K.L., Ding, J., 2012. PAMSCOD: platoon-based arterial multi-modal signal control with online data. *Transp. Res. Part C* 20, 164–184.
- Kloos, C.W., Danaher, A.R., Hunter-Zaworski, K.M., 1994. Bus priority at traffic signals in Portland: the Powell Boulevard pilot project. In: *Compendium of Technical Papers*. Institute of Transportation Engineers.
- Koonce, P., Rodegerdts, L., Lee, K., Quayle, S., Beard, S., Braud, C., Bonneson, J., Tarnoff, P., Urbanik, T., 2008. *Traffic Signal Timing Manual FHWA-HOP-08-024*. Federal Highway Administration, Washington, DC.
- Lämmer, S., Helbing, D., 2008. Self-control of traffic lights and vehicle flows in urban road networks. *J. Stat. Mech: Theory Exp.*, P04019.
- Nash, A., 2001. Implementation of Zürich's Transit Priority Program. *Mineta Transportation Institute, Report 01-13*.
- NCHRP, 2010. *NCHRP Synthesis 403: Adaptive Traffic Control Systems: Domestic and Foreign Practice*. Transportation Research Board.
- Newell, G.F., Osuma, E.E., 1969. Properties of vehicle-actuated signals: II two-way streets. *Transp. Sci.* 3 (2), 99–125.
- Robertson, D.I., Bretherton, R.D., 1991. Optimizing networks of traffic signals in real time – the SCOOT method. *IEEE Trans. Veh. Technol.* 40 (1), 11–15.
- Wadjas, Y., Furth, P.G., 2003. Transit signal priority along arterials using advanced detection. *Transp. Res. Rec.* 1856, 220–230.
- Xie, X., Barlow, G.J., Smith, S.F., Rubinstein, B.Z., 2011. Platoon-based self-scheduling for real-time signal control. In: *Proceedings 14th International IEEE Conference on Intelligent Transportation Systems*, Washington, DC.
- Xie, X., Barlow, G.J., Smith, S.F., 2012. Schedule-driver coordination for real-time traffic network control. In: *Proceedings 22nd International Conference on Automated Planning and Scheduling*, Sao Paulo, Brazil.