Transportation Research Part B 162 (2022) 195–208

Contents lists available at ScienceDirect

Transportation Research Part B

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

Traffic signal plans to decongest street grids

Bassel Sadek *, Jean Doig Godier, Michael J Cassidy, Carlos F Daganzo

Department of Civil & Environmental Engineering, University of California, Berkeley

ARTICLE INFO

Keywords:
Network traffic congestion
Oversaturated network traffic
Network traffic control
Signal coordination

ABSTRACT

Two new synchronization strategies are developed for signalized grids of two-directional streets. Both strategies are found to reduce congestion significantly more than do other approaches. One of the strategies is static and the other adaptive. Both use a common timing pattern for all signals on the grid but use a different offset for each. The static strategy serves the morning rush by providing perfect forward progression on all streets in the directions that point toward a reference intersection, one that is located near the center of gravity of all workplaces. For the evening rush, perfect progression is achieved for all travel directions that point away from the reference intersection. The adaptive strategy toggles between this forward synchronization mode and a second mode suited for congestion, but only in a pre-determined district surrounding the reference intersection. Toggling is based on the district's real-time traffic density.

The paper shows how to switch quickly between the two synchronization modes without resorting to unacceptably short phases. It also shows that if the grid is formed by two intersecting sets of parallel streets, even if unevenly spaced, then every street can be perfectly synchronized in one of its directions. As a result, an inbound driver in the morning, or an outbound driver in the evening, is guaranteed to encounter synchronized signals over the full length of her trip. Although this is not possible for more irregular grids, the paper shows how to modify the two strategies for this case, so that they still perform well.

The strategies were benchmarked with simulations against a fixed, zero-offset strategy for many scenarios, because zero-offsets are known to work well under congestion. In one important scenario representing a severely congested morning rush, both strategies were also benchmarked against a state-of-the-practice computer program. While the state-of-the-practice program reduced the zero-offset delay by a modest 7%, the proposed strategies reduced it by 21% (static) and 32% (adaptive); i.e., improving on the state-of-the-practice program by 14% and 25%. These improvements considerably exceed the 1% to 5% reductions typically reported in the literature for other state-of-the-art methods that have been compared with state-of-the-practice programs. Similarly good results were obtained for the other scenarios, which included the morning and evening rushes, various distributions of workplaces, and both regular and irregular grids.

1. Introduction

There is sizable literature on how best to time in a coordinated way neighboring traffic signals on a street network, e.g., see (Morgan and Little, 1964; Traffic Research Corporation, 1966; Little, 1966; Robertson, 1969; Little et al., 1981; Newell, 1989; Gartner et al., 1991; Gordon, 1996; Hu and Liu, 2013; Yang et al., 2015). The task consists in deciding not just the phasing of each signal, but

* Corresponding author.

https://doi.org/10.1016/j.trb.2022.05.014
Received 11 May 2022; Accepted 27 May 2022
Available online 18 June 2022

also the instants when the green phases for all signals start. For grids run on a common cycle time, C, as in our schemes, the different times when a specific signal turns green for a given direction take on the same numerical value, modulo C, in the [0, C) interval. We call this value the "signal offset"—in contrast to other definitions sometimes used in the literature. The coordination part of the timing plan then consists of selecting the offset for each signal. Offsets are usually chosen to minimize delays, queue lengths or other travel costs (e.g., Gordon, 1996; Robertson, 1969; Traffic Research Corporation, 1966). In the absence of residual queues, the aim may be achieved by synchronizing offsets to maximize the time windows (bandwidths) through which platoons can traverse intersections without stopping (Morgan and Little, 1964; Little, 1966; Little et al., 1981; Gartner et al., 1991; Hu and Liu, 2013; Yang et al., 2015), though bandwidths that are insufficient to accommodate demands can create residual queues and work poorly; see Newell (1989).

When these residual queues grow long, synchronizing offsets to backward waves can be an effective response (Pignataro, 1978; Newell, 1981; Rathi, 1988; Quinn, 1992; Daganzo and Lehe, 2016; Daganzo et al., 2017). This ensures that green phases for a given movement are not initiated until the intersection is clear of any downstream queues that block it, thus increasing the flow that is discharged during the green phase.

As is well known, offsets can only be synchronized in one of the directions of a two-way arterial, so the opposing direction will usually be unsynchronized.[1] To resolve this incompatibility, it is common practice to synchronize offsets only in the direction with heavier flow (Gordon, 1996). More vehicles then benefit from synchronization, which reduces delay.

Additional incompatibilities arise on networks, however. For example, although we can synchronize all N-S streets of a grid in one direction (say from N to S), choosing the offsets at all the intersections of these streets fixes all the offsets across the grid. Therefore, no additional adjustments can be made to guarantee that the intersecting set of streets is also synchronized. Researchers have responded to this challenge by decomposing networks into parts, and separately determining offsets for the intersections on each sub-network (Hillier, 1966; Allsop, 1968; Gartner and Little, 1973; Gartner et al., 1975; Gartner and Stamatiadis, 2002; Gartner and Stamatiadis, 2004; Ye et al., 2015; Liu, 2015). Mathematical models, iterative techniques, indices, and expert opinions have been developed to that end (Ye et al., 2015; Liu, 2015). The focus of all this earlier research is light congestion—i.e., networks without queue spillovers. Simulations under steady state traffic and light congestion show that these state-of-the-art schemes often outperform, albeit only slightly, commercially available computer programs. Reductions in vehicle delay typically range from 1% to 5% (Liu, 2015).

The present paper proposes two simple signal timing strategies to address not just these steady state situations, but also situations with heavy congestion and queue spillovers. The focus is on grids formed by two sets of roughly parallel intersecting streets. Both strategies produce synchronized offsets (forward or backward) in one direction of every link on the network. This direction points toward a reference intersection in the morning rush, and away from that intersection in the evening. Section 2 describes the proposed timing strategies and related supporting facts; Sec. 3 benchmarks the strategies with simulations of multiple scenarios; and Sec. 4 summarizes the results, conveys some caveats, and lists some ideas for future research.

2. Proposed strategies

Considered is an inhomogeneous, rectangular grid formed by two intersecting families of roughly parallel, two-way streets with a traffic signal at every intersection. The proposed strategies time these signals identically, with a common cycle length, C. While these identical timings may be suboptimal for the observed traffic flows at individual intersections, they facilitate synchronization and should be good for the system on a whole. Results will attest to this supposition.

In view of this, the variables needed to define a timing plan for the proposed strategies are C, the common phases and the set of individual offsets. For congested grids, which is the situation of interest, C should be as large as practicable, and this is what we propose.[2] For capacity considerations, we also propose using the smallest possible lost time; and green phases that are inversely proportional to the average number of lanes for the streets in the corresponding directions. For example, if all the streets in the network have the same number of lanes, both the N-S and E-W green phases would be equal. Clearly, given these choices, only the individual offsets remain to be chosen and this shall be our focus.

To this end, the reference intersection shall be the one closest to the Center of Gravity (CoG) of all workplaces.[3] In this way, every link in the network will be synchronized in the direction that is likely to carry most traffic—pointing toward the CoG in the morning and away from it in the evening. Two such directional links are notated on Fig 1, where the reference intersection is 5.

For simplicity of exposition, we will initially assume that our two street families (i.e., N-S and E-W) are perfectly parallel as in Fig 1, i.e., that they display translational symmetry with the same but uneven block lengths. We will address irregular grids with non-parallel streets in later sections. We will also assume, again for simplicity, that the same fundamental diagram holds for all links in the network; and consequently, that signals are timed with two green phases of equal duration. These assumptions can also be relaxed, as we will discuss in Sec. 4.

The present section describes the proposed strategies, and their properties. Subsections 2.1–2.3 address the morning rush, presenting a static strategy (for undersaturated traffic) in Sec. 2.1; a static strategy (for oversaturated traffic) in Sec. 2.2; and an adaptive strategy (for all traffic types) in Sec. 2.3. Section 2.4 then explains how to modify these schemes for the evening rush. The strategies are finally generalized in two ways: to networks with widely separated workplace clusters in Sec. 2.5, and to irregular grids in Sec 2.6.

[1] An exception occurs if the time it takes a vehicle/wave to travel the block is a multiple of $C/2$.
[2] Cities commonly adopt policies concerning acceptable maximums for cycle length and green times.
[3] The CoG is the average workplace location, weighted by the number of workers at each work location.

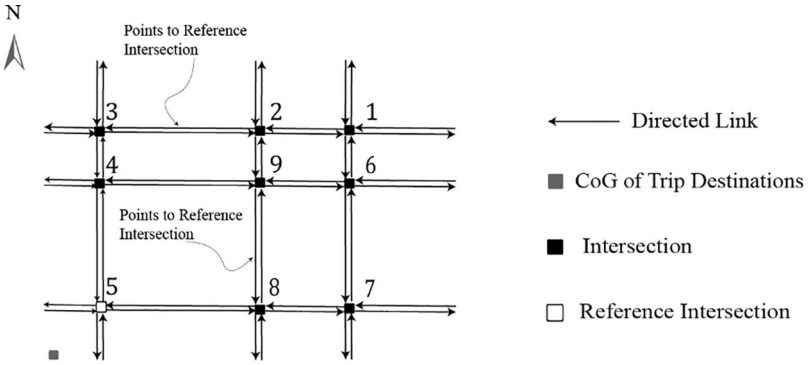

Fig 1. 3 × 3 Example Network.

2.1 Undersaturated traffic in the morning rush: focused forward progression

A static strategy for these conditions is now presented. The idea is to achieve forward progression toward the reference intersection on every block of the grid. This section will prove that a driver will enjoy synchronized signals while traveling along any path that points toward the reference, e.g., intersection 5 in Fig 1. Since the reference intersection is the focus of all progressions, the method will be called "Focused Forward Progression" or FFP. Details follow.

It will be assumed without loss of generality that the two street families are oriented in the N-S and E-W directions.[4] Then, we define the offset of every signal as the time t (modulo C) when the signal's E-W green phases start. However, because all signals are identically timed, offsets can also be defined with the same results with reference to any point in the signals' cycles, rather than at the beginning of the E-W green phase.

To define offsets for the proposed mode of synchronization, let x_j and y_j be the distances along the E-W and N-S directions separating an intersection j and the reference intersection. We take these values to be positive regardless of the direction from j to the reference intersection. The proposed offsets are then:

$$\delta_j = \left(\frac{x_j + y_j}{-v_f} \right) \text{ (modulo C)} \tag{1}$$

where v_f is the free-flow vehicle speed. This ensures that an observer traveling at the free flow speed along any shortest path from j to the reference intersection would see the same point in the cycle at all intersections visited, e.g., the beginning of the E-W (or N-S) green phases.

We shall say that a directed link of our grid is "synchronized" (with FFP) if, according to convention, an observer traveling along the link would see the same point in the cycle at both ends of the link. We shall also say that a link "points" to the reference intersection if the link's "to" node is closer to the reference intersection than is the link's "from" node. Note that 50% of the links in a grid point to the reference intersection, and the other half point away from it.

We can now state the following proposition.

Proposition 1. Under (1), every link pointing to the reference intersection is synchronized.[5]

Proof: Since the network is formed by two sets of parallel streets, every link pointing to the reference intersection is on a shortest path to it. Under (1), an observer traveling at the free flow speed along this shortest path sees the same point in the cycle at all intersections visited. Therefore, the observer sees the same point in the cycle at the two signals at the ends of our link. Therefore, the link is synchronized. □

This proposition implies that the FFP method synchronizes every street block with forward progression in the direction that points toward the reference intersection, as was claimed. Thus, a single vehicle can travel on any street toward the reference intersection without interruption, except when turning.[6] Note that FFP can be restricted to any geographical region one wishes by using (1) only for the nodes contained in the region. In this case, Proposition 1 continues to hold but only for the links entirely contained in the select region.

[4] All the results that follow apply equally if the orientation of the grid is different, and even if the two sets of streets intersect at an oblique angle.

[5] Although we have used rectangular grid notation to this point, this is only for ease of presentation. Consideration shows that the proposition also applies to non-rectangular grids (e.g., slanted, even with curvy streets) provided they exhibit our form of translational symmetry.

[6] By turning at intersections, vehicles change the direction in which green phases are needed, so vehicles start the trip on their next link under a red. Since the link is synchronized, they meet a red phase at the end, experiencing a one-time penalty. Because of this turning penalty, commuters traveling in uncongested conditions should favor paths with only one turn.

2.2. Oversaturated traffic in the morning rush: focused backward progression

If traffic is oversaturated with queue spillovers, commuters should still predominantly use links that point to the reference intersection. For this reason, our static strategy for oversaturated traffic will continue to focus on these links. The only difference from FFP is that offsets shall now be synchronized with the backward waves of kinematic wave theory. This strategy can be applied to the whole grid or to any district within the region containing multiple links. Called "Focused Backward Progression" or FBP, the scheme will be shown to synchronize in the desired way every link that points to the reference intersection in the district where it is applied.

Under FBP, offsets in the congested district are set using the shortest paths toward the focus (i.e., reference intersection) but using an observer who travels with the backward waves at speed $-w$. Using primes to refer to the backward mode, the formula is:

$$\delta_j^{'} = \left(\frac{x_j + y_j}{w} \right) \text{(modulo C)} \tag{2}$$

In this expression x_j and y_j continue to be the (always positive) distances separating j and the reference intersection along the E-W and N-S directions.

With this formula, an observer traveling along a shortest path of directed links pointing to the reference intersection, but in the reverse direction of traffic (i.e., backwards, away from the reference) and at speed w, will see all the signals on its path at the same point in their cycles. Obviously then, a modified version of Proposition 1 continues to hold if we just use the relevant observer, i.e., we have:

Proposition 2. Under (2), all links that point to the reference intersection are synchronized.

Therefore, the FBP recipe (2) synchronizes all blocks in the direction pointing toward the reference intersection with backward progression.[7] As with FFP, the FBP mode can be restricted to any region one wishes, so that Proposition 2 continues to hold for all the links entirely contained in said region.

2.3. Adaptive strategy for the morning rush: toggling between synchronization modes

The proposed adaptive strategy starts with the FFP mode (1) everywhere, and then adaptively toggles between FFP (1) and FBP (2) inside a predefined district. This district is centered on the CoG and is optimally sized. Toggling depends on the average traffic density in the district, as explained below. Outside this district, the FFP mode (1) is used all the time.[8]

In the following we will say that a signal is "properly set" in a time interval for a given mode if all mode-specific observers passing through the signal at all times during the interval see the same signal colors at the signal in question and at the reference intersection. Note that when this happens for all the signals of a street then, starting at the beginning of the interval, the street's time-space diagram will exhibit "green bands" with the progression speed of the observer that are as wide as possible, i.e., terminating when either a signal turns red or beyond the interval.

Keeping this in mind, we see that if we instantaneously toggle from (1) to (2), or vice versa, at some decision instant, t_s, then the signals will be properly set for the desired mode at all times before and after t_s. Therefore, every street should display wide green bands with the correct speed on both sides of the line $t = t_s$. This is illustrated by Fig 2(a), where the switch is from (1) to (2), and the reference intersection is beneath intersection 7. Note the wide green bands in the figure.

Note too that as a result of this instantaneous toggle, the signal colors on each side of $t = t_s$ may not coincide at some intersections, e.g., intersections 5–7 in Fig 2(a). When this happens, a phase change is needed at $t = t_s$ and some very short phases may result. A phase change is not necessary if the colors coincide, as occurs for intersections 1–4 in the figure. In this case, however, very short phases can still occur, as can be seen for intersection 2. Since phases of abnormally short length cannot be used in real life, e.g., for pedestrian safety,[9] we shall introduce a minimum required length, φ, for all phases. Phases labeled I, II, III, IV in the figure violate this constraint, so the quick transition is infeasible for our example.

Short phases can be removed in many ways. The Appendix describes two methods. The first of these has two steps. The first step consists of reversing the color of a single, carefully selected phase at each intersection that has an unfeasibly short phase. This always succeeds in removing these phases but can create unreasonably long ones. The second step eliminates the longest of these by inserting in a systematic way one or more phases of opposite color in the long phase's midst. Fig 2(b) shows the outcome for our example. It shows that the method eliminates all short phases, properly sets all the signals within a single cycle, and keeps the duration of all abnormal phases under 3φ, which is reasonable. The appendix describes the details.

The second method has multiple steps and allows the user to specify tighter upper limits to the phase lengths. A third algorithm is described in (Sadek, 2021). In our experience, the results of the simulations are not significantly affected by the choice of the transition algorithm. The choice thus becomes a matter of taste.

2.4. The evening commute: dispersing progression

The ideas of the previous sections extend trivially to the evening commute by reversing directions. The shortest paths and directed

[7] Curiously, and unlike in the forward synchronization case, vehicles now save time when turning if the network is congested; see (Sadek, 2021).
[8] When the district is in FBP mode, links crossing the district's boundary may be unsynchronized.
[9] Minimum durations are often set to accommodate pedestrian crossings, e.g., (Federal Highway Administration, 2005)

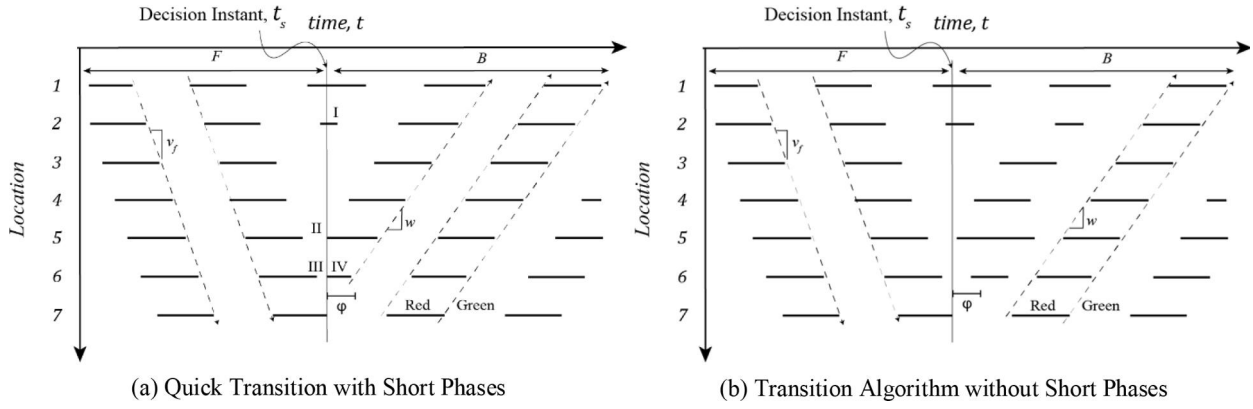

Fig 2. Switching Between Coordination Strategies.

links of sections 2.1 and 2.2 now should fan away from the reference intersection, rather than point toward it, to disperse evening traffic more easily. Moreover, the imaginary observers should now travel in directions opposite to those of the morning rush. The only substantive change in the new procedures (which we call DFP for "Dispersing Forward Progression" and DBP for "Dispersing Backward Progression") is that the denominators of (1) and (2) change in sign. The rule for toggling between modes in a district and the transition algorithm stay the same. As the reader can surmise, both DFP and DBP achieve synchronization on 50% of all links, i.e., those that fan away from the reference intersection.

2.5. *Generalization for widely separated workplace clusters*

If the workplaces are dispersed into widely separate clusters, it may be of benefit to decompose the network into parts associated with each cluster and treat these parts separately. To apply the idea, first partition the workplaces into non-overlapping clusters, and define a CoG for each. Then assign each intersection to its closest CoG to form a subnetwork. Finally, define the offsets for the intersections in each of these subnetworks with either the FFP or the adaptive method, independently of the other subnetworks. An illustration will be furnished in Sec 3.3.

2.6. *Geometrically irregular grids*

A real-world grid may lack symmetry, e.g., because its links are somewhat curved with irregular lengths or are not perfectly parallel. In these common cases, (1) and (2) may produce poor synchronization for some links, creating a cascade of negative consequences. To alleviate this problem, we propose modifying the recipes so that they will produce reasonably good synchronization on all the links, rather than perfect synchronization on some and very poor on others.

To do this, an auxiliary rectangular grid of the type used in previous sections and topologically equivalent to our irregular network is defined. The common length of the links in any column or row of this auxiliary grid is set equal to the average of the lengths of the corresponding links in the original irregular network. Formulas (1) and (2), or their correspondents for the evening commute, are then applied to the auxiliary rectangular grid. The offsets found for each node of the auxiliary grid are then applied to the topological corresponding nodes of the original grid. The transition algorithm is used without any changes because it does not use link lengths. Section 3.5 demonstrates and tests the idea.

3. Numerical comparisons

The proposed timing plans were simulated on an inhomogeneous grid formed by two perpendicular sets of 20 parallel, but unevenly spaced streets with two lanes in each travel direction. The AIMSUN simulation platform (Aimsun, 2018) was used.[10] The separation between streets ranged from 150 m to 250 m, commensurate with the block lengths of many cities (Siksna, 1997). Fig. 3 shows the complete network.[11] The district where offsets are adaptive was chosen to be the highlighted 6 × 6 region in the network's center, except when stated otherwise.[12] All the signals were identically timed with C = 90 s and two equal phases, each with 44 s of effective green. The fundamental diagram was the same everywhere with v_f = 50 km/h, w = 18 km/h, and optimal density, k_0 = 45 veh/km/lane.

The human side was simulated as follows. Trips were generated at the rates shown by the cumulative vehicle count curve in Fig 4

[10] Its default settings were used except as otherwise noted.
[11] Two deformed versions of this network were also tested, although not as extensively.
[12] This district size of 6x6 worked best for most of the scenarios tested.

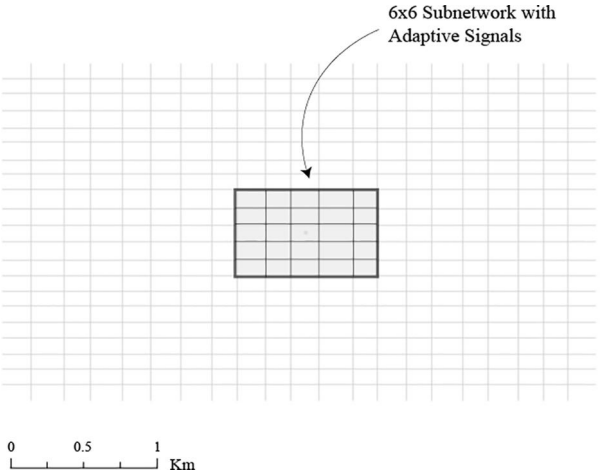

Fig 3. 20 × 20 Grid Network.

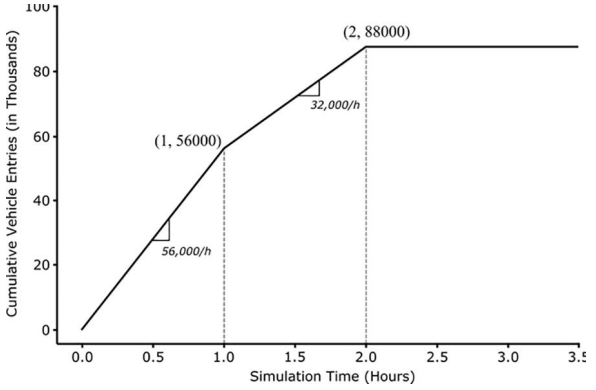

Fig 4. Cumulative Vehicle Entries into the Network.

and were equally divided across every home-workplace pair. Homes (i.e., trip origins in the morning and destinations in the evening) were uniformly distributed over the network. Workplaces were distributed in various ways. Drivers were assumed to navigate adaptively, based upon average travel times on the links. For a measure of realism, these travel times were updated at 6-min intervals and were delivered to 30% of the drivers on the network, who were randomly sampled each interval. These drivers then revised their routes, as per the AIMSUN logic (Cascetta et al., 1996).[13] The set-up severely congested the network and illustrations of this follow.

Results were encouraging. Section 3.1 describes them for the morning commute, assuming workplaces were somewhat concentrated at the center of the network. The remaining sections examine variants of this basic scenario. Sections 3.2 and 3.3 consider different workplace distributions: dispersed in Sec. 3.2, and multimodal in Sec. 3.3. Section 3.4 then tests the evening commute, and Sec. 3.5 two deformed grids. Except where otherwise stated, all outcomes presented are averages of 10 simulations with different random seeds.

3.1. Basic scenario: morning commute with concentrated workplaces

We use the network shown in Fig 3. Workplaces were randomly generated with a Gaussian distribution centered on the CoG, and a standard deviation such that 40% lied in the 6 × 6 central district. The critical density for toggling between synchronization modes in this district was 45 vehicles/km/lane, which is the lowest density yielding capacity flows.[14]

[13] The logic uses link travel times at each 6-min interval as inputs to a logit model. It, in turn, has drivers select the links to be used for the remaining portions of their trips, to minimize individual travel times. The 30% of drivers who received traffic updates was AIMSUN's default value. We suspect that this default was calibrated to real-world conditions. Our own parametric analyses found that using lower percentages had negligible effect on outcomes, while higher percentages produced aberrant driver behavior. More will be said about adaptive routing in Sec. 4.

[14] This is the average density at which the district's exit function began to plateau, as observed in AIMSUN simulations. The inspection interval for identifying over-saturated conditions was set at 6 min, an integer multiple of the system's 90s cycle length.

Simulations of the FFP and adaptive strategies were performed along with two commonly used fixed-time strategies: the zero-offset benchmark and a state-of-the-practice program called SYNCHRO (Trafficware, 2012).[15] The network became severely congested under each of these strategies, but to varying degrees.

For illustration, Fig 5 shows the sample path of vehicle accumulation vs trip-completion rate for a single simulation of the zero-offset benchmark scenario. The numeric labels along the curve are time stamps in minutes. Note the high accumulation from minute 60 to 120, while vehicles are still entering the network, which is nearly double the 7500 vehicle "optimum accumulation" at $t = 25$. For comparison, the figure also displays the system's state at $t = 120$ (the end of loading) for the other three strategies; see the black triangle and two circles. Note how the proposed adaptive strategy reduces accumulation by about 40% from the zero-offset benchmark, and in so doing decongests the network considerably. As a graphic illustration of this benefit, Figs. 6(a)−6(d) show, for each strategy, the congested footprint of the network at $t = 120$. Note the gradual but significant reduction in size from (a) to (d). As further illustration, Figs. 7(a)−7(d) show the same gradual pattern 30 min later, when the zero-offset strategy still exhibits a significant footprint, but adaptive strategy (d) has already eliminated the footprint, totally decongesting the network. In the simulations, the three other strategies required at least 10 more minutes to eliminate the footprint.

Finally, the time-series of simulated vehicle accumulations are presented for each of the four strategies in Fig 8. Recall that the area under each of these curves is the total vehicle-hours traveled (VHT). The upper curve reveals again that congestion was most severe under the zero-offset benchmark, as it yielded 21,755 h of VHD (vehicle-hours of delay, including both signal delay and congestion delay). This amount corresponds to about 13 min of delay per vehicle. Note how the curves are much lower for the proposed strategies, signifying considerable reductions in VHT relative to the zero-offset baseline and the plan generated by SYNCHRO. Reductions in VHD, which are more relevant because VHD excludes the common free-flow travel time that cannot be changed, are not shown. Instead, they are summarized in Table 1 alongside the VHT reductions. Note how the adaptive strategy is spectacularly effective, reducing VHD by 32% – about five times more than what SYNCHRO achieved – thus erasing 4 min of delay from the average trip. Even the simple FFP strategy turns out to be three times more effective than SYNCHRO. These outperformances are noteworthy because previously proposed strategies outperformed SYNCHRO by only 1% to 5% (Liu, 2015).

We suspect that the more dramatic improvements offered by the proposed methods occurred because SYNCHRO optimizes green splits at each signal on the network, based upon local conditions. This makes it difficult, if not impossible, to synchronize offsets network wide. The finding underscores the value of timing green splits (and cycle length) identically across the network, as we have done, to facilitate coordination for all inbound or outbound trips. Given this finding, and the relative difficulties in using SYNCHRO, we will no longer use SYNCHRO as a benchmark in the analyses that follow.

3.2. Morning commute with dispersed workplaces

We repeated the experiment of the previous subsection in settings with the same global demand rates but more dispersed workplaces—with six dispersion levels in total. Table 2 shows the results. For each dispersion level, the table gives the average VHD with the zero-offset baseline, and the two proposed strategies. Percent reductions that the latter achieved from this benchmark are shown as well. The scenario of Sec. 3.1, with the greatest concentration of workplaces, is placed across the top of the table as the point of reference. The extreme (and very unlikely) scenario with a perfectly even distribution of workplaces over the network is placed at the table's bottom.

As one might expect and as the table shows, baseline congestion for the zero-offset strategy declined steadily from top to bottom, i.e., with the level of dispersion, because fewer trips then need to use the central part of the network where the district resides. In particular, the middle two scenarios exhibit VHDs corresponding to about 8 and 6 min of delay per car, signifying considerable but not enormous congestion.

The table also shows that the percent improvement in VHD for the adaptive FFP+FBP strategy was lower for the lightly congested scenarios near the bottom of the table—albeit still in double digits, except for the unrealistic case at the bottom. Such a decline in performance makes sense because low congestion implies few queue spillovers that adaptive offsets can mitigate.

Curiously, a similar declining pattern is observed with the static FFP strategy. The reason is likely to be different, though. With increasing dispersion, many trips do not go toward the network's center. As a result, we surmise that these trips may be slowed by the centripetal progression of FFP—thus diluting the strategy's benefit

3.3. Multimodal distribution of workplaces

This scenario differs from the basic scenario of Sec. 3.1 in the location of workplaces, which were distributed by superposing two

[15] The turning ratios required by SYNCHRO were determined for each intersection heuristically. A provisional set of ratios was first obtained by running a simulation with zero-offsets. SYNCHRO was then run with these provisional ratios to obtain a tentative set of offsets. The turning ratios used as inputs for our tests were then obtained by running a second simulation with the tentative offsets. Network-wide vehicle hours of delay under each set of offsets differed by less than 1%, and for this reason further iterations to refine offsets were not performed. Strategies that adapt signal-timings in response to driver routing decisions (e.g. Varaiya, 2013; Hu et al., 2013; Gregoire et al., 2014) were not used for comparisons in this work. This is because timing signals to the routes that drivers want to take is known to be problematic on networks where drivers can choose their routes. Accommodating drivers in this fashion is almost never system-optimal and misses the opportunity to use the signals to encourage efficient routing; see (Smith, 1979) for additional discussion and counterexamples.

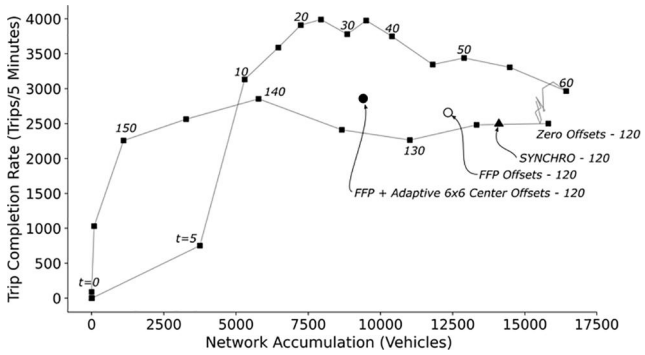

Fig 5. Sample Path for Zero Offsets; and data points at 120 min for alternative strategies.

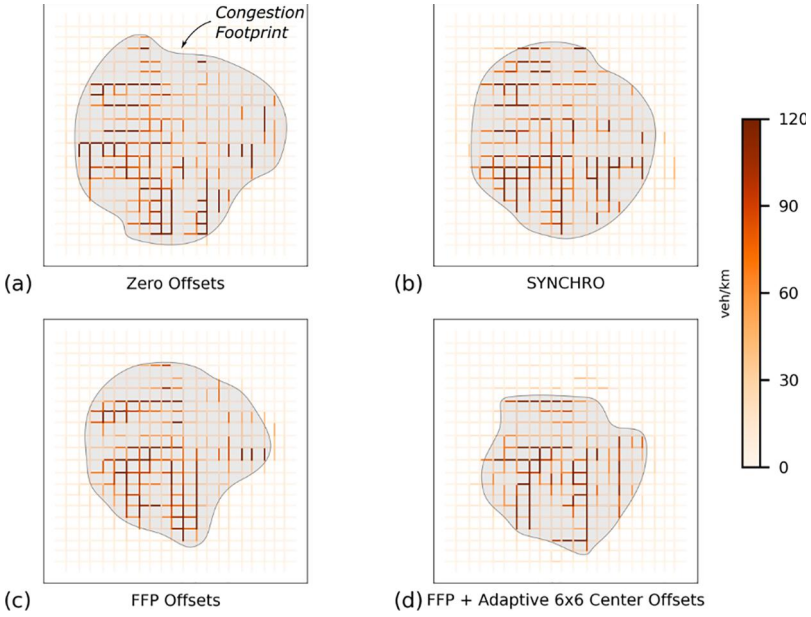

Fig 6. Link Density Plots at $t = 120$ min. (a) Zero Offsets, (b) SYNCHRO, (c) FFP Offsets, (d) FFP + Adaptive 6 × 6 Center Offsets.

identical Gaussian clusters centered on symmetric locations; see Fig 9. The shading intensity of the dots in the figure represents the number of workplaces around each location. Note the two poles with maximum concentration. As before, trips were equally likely to go from any home to any workplace.

The scenario was tested for five offset-setting strategies: the benchmark; the FFP and adaptive strategies in which the CoG resides at the mean location of all workplaces, i.e., at the geometric center of the grid; and the generalized FFP and adaptive strategies of Sec. 2.5 with two CoGs. For the latter case, two non-overlapping 6 × 6 districts centered on each CoG were used for toggling between FFP and FBP. Table 3 shows the results.

Note how the baseline, zero-offset strategy produced 12,964 h of VHD, or about 7 min of delay per car. This is comparable to the figures for the two middle scenarios of Table 2, but less than the baseline figure at the top. Such a reduction in baseline congestion is logical because, as Fig 9 illustrates, workplaces are now more dispersed than in the concentrated scenario at the top of Table 2.

Note as well how in the current case the static and adaptive strategies with one CoG produce VHD reductions like those of the two comparable scenarios in the middle of Table 2. This suggests that for a given level of dispersion, the particular shape of the workplace distribution is not very important, i.e., the proposed strategies are robust with respect to this distribution. Finally, note from the table how the strategies with two CoGs do not yield significant additional improvements compared with one CoG.[16] This failure might be unexpected,[17] but suggests that the basic strategy with one CoG is quite robust.

[16] The adaptive strategy with two CoGs underperforms the one with one CoG. This probably occurs because the total area in which adaptation is applied (which consists of two 6x6 squares) is too large.

[17] The two-CoG strategy will surely improve matters in situations where the two poles of demand are further apart, but this scenario was not tested.

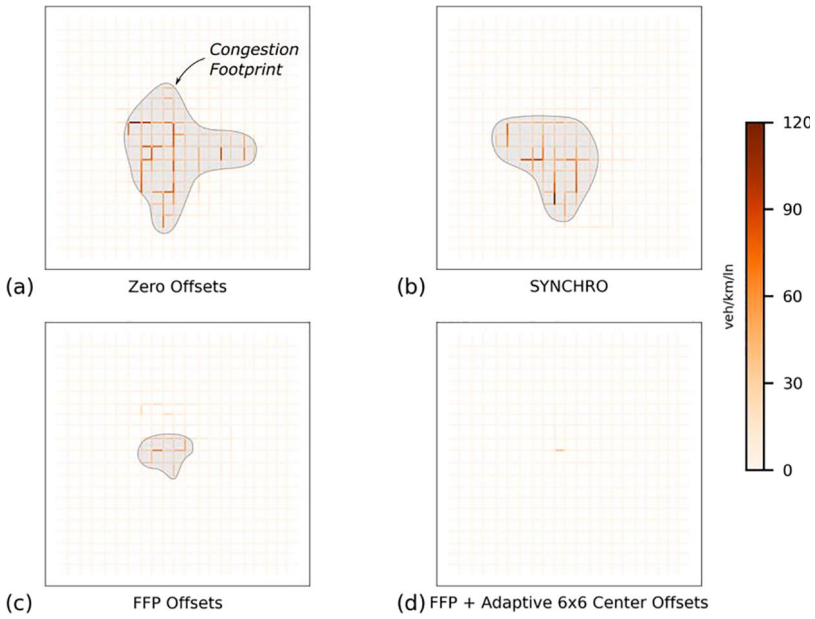

Fig 7. Link Density Plots at $t = 150$ min. (a) Zero Offsets, (b) SYNCHRO, (c) FFP Offsets, (d) FFP + Adaptive 6 × 6 Center Offsets.

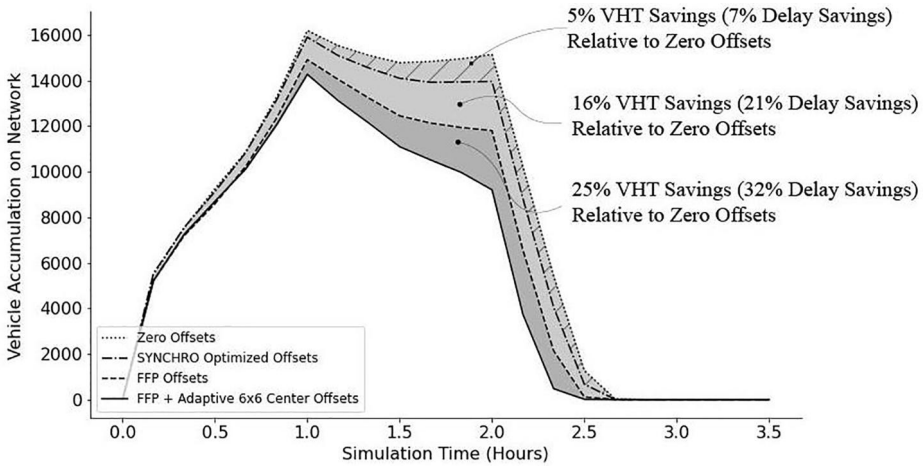

Fig 8. Vehicle Accumulation on the Network for Different Coordination Strategies.

Table 1
VHT and VHD Results for the Morning Commute with Concentrated Workplaces.

Strategy	VHT (in hrs)	VHT Improvement (%)	VHD (in hrs)	VHD Improvement (%)
Zero Offsets	28,318	–	21,755	–
SYNCHRO	26,890	5.0	20,328	6.6
FFP	23,837	15.8	17,273	20.6
FFP + FBP (Adaptive)	21,346	24.6	14,785	32.0

3.4. Evening commute

The scenario of Sec 3.1 was also simulated for the evening commute, simply by designating workplaces as origins and homes as destinations while keeping everything else the same. Table 4 shows the results. It includes an additional scenario with an enlarged 8 × 8 central district because the enlargement reduced VHD even more. Note how the baseline, zero-offset strategy results in less congestion than in the morning. Visual inspection of the simulations suggests that this occurs because queues in this case stay more

Table 2
VHD Results for the Morning Commute with Dispersed Workplaces.

Dispersion Level	Strategy	VHD (in hrs)	VHD Improvement (%)
40% of Trips Destined to 6 × 6 Center	Zero Offsets	21,755	–
	FFP	17,273	20.6
	FFP + FBP (Adaptive)	14,785	32.0
35% of Trips Destined to 6 × 6 Center	Zero Offsets	18,020	–
	FFP	12,671	29.7
	FFP + FBP (Adaptive)	11,055	38.6
30% of Trips Destined to 6 × 6 Center	Zero Offsets	13,898	–
	FFP	11,371	18.2
	FFP + FBP (Adaptive)	9727	30.0
25% of Trips Destined to 6 × 6 Center	Zero Offsets	10,013	–
	FFP	9224	7.9
	FFP + FBP (Adaptive)	8679	13.3
20% of Trips Destined to 6 × 6 Center	Zero Offsets	9645	–
	FFP	8985	6.8
	FFP + FBP (Adaptive)	8551	11.3
9% of Trips Destined to 6 × 6 Center (Uniform)	Zero Offsets	7958	–
	FFP	8137	−2.2
	FFP + FBP (Adaptive)	7906	0.6

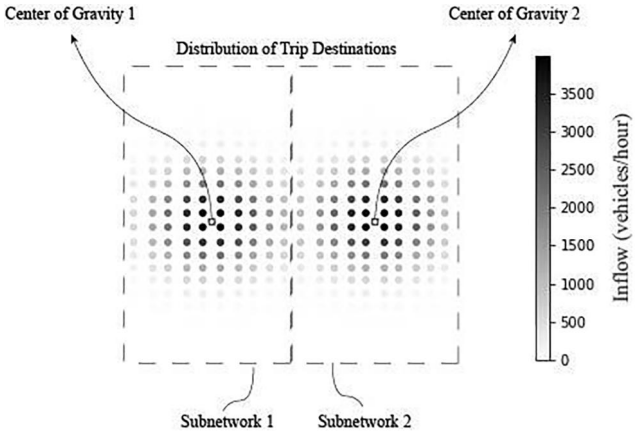

Fig 9. Multi-centric Demand Distribution and Subnetworks.

Table 3
VHD Results for the Morning Commute with Multimodal Distribution of Workplaces.

Strategy	VHD (in hrs)	VHD Improvement (%)
Zero Offsets	12,964	–
FFP (One CoG)	10,757	17.0
FFP + FBP (Adaptive - One CoG)	9954	23.2
FFP (Two CoGs)	10,517	18.9
FFP + FBP (Adaptive -Two CoGs)	10,377	20.0

Table 4
VHD Results for the Evening Commute with Concentrated Workplaces.

Strategy	VHD (in hrs)	VHD Improvement (%)
Zero Offsets	12,843	–
DFP	9637	25.0
DFP + DBP (Adaptive)	9221	28.2
DFP + DBP (Adaptive - 8 × 8)	8527	33.6

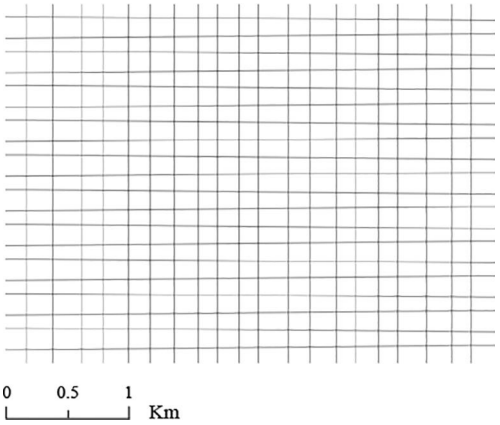

Fig 10. Extreme Asymmetric Grid.

tightly confined within the central district. Curiously, despite the lessened baseline VHD, the proposed strategies reduced it in percentage terms surprisingly well—just as much as in the morning.

3.5. *Irregular networks*

The simulations of Sec. 3.1 were again repeated with two irregular 20 × 20 grids. They differed from the ideal only in the individual link lengths, as explained below.

The first of these grids is shown to scale in Fig 10. The only difference from the grid of Sec. 3.1 is that the horizontal streets are now slanted at alternating angles of ± 0.725° from the horizontal; and that to eliminate vertical links of especially short lengths, the average separation between each pair of horizontal streets was increased by 50 m. As a result, the vertical links on each horizontal row now differ in length at the E and W ends by a large 100 m. Because with the proposed strategy all links in each row share the same relative offsets, no row can be synchronized well. The extreme form of irregularity for this grid is unrealistic but was used to stress-test the proposed strategies to the limit.

The second grid was chosen more realistically, simply by adding a zero-mean random length to each link. The standard deviation was commensurate with that of downtown Los Angeles (California); inspection of links obtained from Open Streets Map (Boeing, 2017) show that the standard deviation is approximately 20 m.

Table 5 shows the results for the basic grid and the two irregular grids side by side.[18] For the extreme scenario, and as one might expect, the static FFP strategy failed to reduce VHD. However, adaptation produced a double-digit improvement. For the more realistic irregular case in particular, the improvements were noteworthy, suggesting that the proposed approach can work well in real-world settings.

4. Conclusions

The paper has presented two new ways of timing the signals of street grids, so as to decongest them more effectively. Both methods apply a common timing pattern to all the signals, using the longest practicable phase lengths and cycle time. Thus, the phases of every signal are predetermined and only the offsets are optimized. Two offset patterns were proposed: one static and the other adaptive. The adaptive solution toggles back and forth between two static patterns, based on the average traffic density measured in a certain district. The two proposed strategies are different from others because those proposed are not responsive to disaggregate traffic conditions. In fact, the static strategy is completely independent of traffic flows, as it simply synchronizes every street block of the network in a centripetal direction during the morning rush and in a centrifugal direction in the evening. The adaptive strategy only depends on the average density measured in a select district. Perhaps the proposed methods perform so well because rather than adapting to local traffic, they force traffic to adapt to the controls, which are then set to facilitate smooth flow, as theoreticians have long recommended.

Both strategies considerably outperformed the zero-offset strategy that is often recommended for congested traffic (Caltrans, 2020), as well as the state-of-the-practice program SYNCHRO, which times network signals to accommodate local link flows, as is the convention. The adaptive strategy reduced delay beyond what could be achieved by SYNCHRO by more than 25%. This is to be compared with other strategies in the literature, which typically reduce delay by 5% or less. The static strategy also performed well, and analyses of both strategies spanned a range of conditions including different congestion levels, concentrated, dispersed and poly-clustered workplaces, as well as the traffic patterns of the morning and evening commutes. The proposed strategies also performed remarkably well on an irregular grid. Our success in this latter case was achieved simply by using suitable averages, and suggests that other commonplace complications of real-world networks, such as fundamental diagrams that are distinct across links,

[18] Because the extreme grid was enlarged, it had more room to hold vehicles and exhibited less congestion.

Table 5
VHD Results for the Basic and Asymmetric Grids.

Strategy	VHD Basic Grid	Improvement (%)	VHD Extreme Asymmetric Grid	Improvement (%)	VHD Realistic Asymmetric Grid	Improvement (%)
Zero Offsets	21,755	–	13,497	–	35,694	–
FFP	17,273	20.6	13,577	−0.6	32,294	9.5
FFP + FBP (Adaptive)	14,785	32.0	11,338	16.0	27,716	22.4

can also be addressed in similar fashion.

There are some caveats to all this promise, however. The results are based on simulations, which may not be perfectly realistic. In the authors' opinion, the Achilles heel of any network simulation, including ours, is the adaptive routing used for the drivers. This is a source of concern for us because drivers are idiosyncratic, difficult to model and we have seen in our simulations that the results depend on what is assumed more than we expected.[19] Therefore, despite the promising results reported in this paper, final say should rest on field tests. The static FFP strategy is relatively simple to implement, and we believe should be tried in the field first.

Another caveat is that the proposed signal timing methods have been tested only on completely signalized grids, which is an uncommon setting in the real world. Although the proposed methods can be applied without modification to partially signalized grids, there is reason to suspect performance would suffer in these cases. More tests and perhaps modifications to the methods would be desirable.

Yet another caveat is that few networks in the real world are composed of only 4-legged intersections. Fortunately, in many of these cases (e.g., think of Manhattan) one may be able to identify an underlying street grid by ignoring just a few streets and links. When this happens, could the proposed strategies be applied to the underlying grid and in so doing improve performance of the entire network? This seems another interesting research avenue to pursue.

One final caveat is that we have not tested all possible network geometries and demand levels, e.g., with ultrashort links, larger networks, and even more diverse levels of congestion. Therefore, additional tests and case studies could be of interest.

Also of interest are extensions that would account for more complex signal timings, say with the inclusion of protected turn phases. Future work might also explore toggling between forward- and backward synchronization on a link-by-link basis based on localized traffic information, even though this may negate the benefit of forcing drivers into corridors with a desired type of progression, as the current strategies seem to do.

Acknowledgments

The research was funded by the National Science Foundation Grant No. 1760971 and by the National Institute for Congestion Reduction (NICR), a University Transportation Center. We thank Associate Editor Michael Zhang for his helpful comments in readying the paper for publication.

Appendix: The Transition Algorithms

Before starting, it will be convenient to introduce some notational artifacts that will simplify the discussion. First, assume without loss of generality that the origin of time is placed at the decision instant, so $t_s = 0$. Second, recall from Fig 4(a) that when we toggle instantaneously, some phases straddle the decision instant. Because the existence of such phases complicates the narrative, we shall artificially divide each of them into two consecutive phases of the same color, i.e., by imagining that there is a phase change at $t = 0$ that does not change a phase's color. In this way all signals experience a phase change (either virtual or real) at the decision instant. All the divided phases, and any others that happen to begin or end at $t = 0$ will be called from now on "boundary phases." Note that by construction, every signal now has two boundary phases—one before and one after $t = 0$.

Algorithm 1: The algorithm has two steps. The first removes all infeasible short phases, and the second (which is optional) shortens some of the long composite phases that result.

Definitions: After obtaining the initial pattern with the instantaneous toggle, we number the phases (including those resulting from divisions) at each intersection consecutively, 1, 2, 3… increasing with time, and starting with the boundary phase preceding $t = 0$. We also use t_i for the time when the i^{th} phase begins. The first step is as follows.

Step 1: For each traffic signal with a need for a phase adjustment, reverse the color of phase 2. If this doesn't remove the signal's infeasibilities, reverse instead the color of phase 3.

Proof it works: Infeasibilities occur only in two cases. If phases 1 and 2 are of different color and either, or both, have a duration of less than the minimum, φ. Or, if the combination of the short same-color phases 1 and 2 is less than φ. Flipping the color of phase 2 will always eliminate infeasibilities of the first case because after the flip phases 1, 2 and 3 will be of the same color, and phase 3 is known to be feasible. In the second case, flipping the color of phase 3 equates the color of phases 1 through 4, so they become a single long (and

[19] Of particular concern is the route choice model embedded in AIMSUN, which is not realistic when, as occurs in our case, there are many routes with complex overlapping patterns (Cascetta, 2001).

feasible) composite phase.

Notes: Normal phases begin either at $t_5 < \varphi + C$ if phase 3 is flipped, or at time $t_4 < C$ if phase 2 is flipped. Very long phases can occur, e.g., if phase 3 is flipped the composite phase can be longer than a cycle. The second step trims some of these long phases by inserting into them a phase of the opposite color, provided that such insertion does not create infeasibilities.

Definitions: The time interval of the target phase is denoted $[t_B, t_E]$, and that of the inserted phase (x*, y*). The step is as follows.

Step 2: Calculate $x^* = \max\{t_B + \varphi, t_s\}$ and $y^* = (t_E - \varphi)$. Then, if $y^* - x^* \geq \varphi$ reverse the color of the signal in the interval (x*, y*) and repeat the procedure with the inserted phase as the target until the problem becomes infeasible.

Proof it works: The end points x and y of an inserted interval of opposite color can be chosen by maximizing the interval's length subject to feasibility constraints, ensuring that the three resulting phases are of sufficient duration, and that the insertion does not start before t_s; i.e., by solving the following LP: $\max\{(y-x)$ s.t.: $(x - t_B), (t_E - y) \geq \varphi; (y - x) \geq \varphi; x \geq 0\}$.

By relaxing the constraint involving the middle interval, the problem decomposes, and we can solve for x and y separately to find: $x^* = \max\{t_B + \varphi, 0\}$; $y^* = t_E - \varphi$. This describes a feasible insertion if $y^* - x^* \geq \varphi$, in which case the interval is inserted. Otherwise, no insertion occurs. This is what is done in the step.

Notes: Consideration shows that the longest phase remaining after step 2 cannot exceed 3φ. Also note that the step 2 procedure does not change the time when the normal phases begin, and that full synchronization begins before that time – with the start of the last abnormal phase immediately before. In the worst case, when the color of phase 3 is changed, normal phases begin at time $t_5 < \varphi + C$. Furthermore, since the last abnormal (but feasible) phase must be greater than φ, it follows that synchronization starts no later than $t = C$. This inequality also holds when phase 2 is flipped, as then the first normal phase begins at $t_4 < C$.

Algorithm 2: This algorithm eliminates infeasible phases while capping the duration of long phases at $T_i + e_i$ time units, where Ti is the original duration of the i^{th} phase and e_i is a chosen parameter. Virtual switches and phases are not used now. The real phases are numbered consecutively starting with either the phase straddling the decision instant $t = 0$ if there is a straddling phase, or else the first boundary phase. We use φ_i to denote the minimum required phase length for the i^{th} phase (which depends on its color).

Recursive step: For each intersection, reset t_i to a new value t_i' with the recursion:

$t_1' = t_1$; $t_{i+1}' = \text{middle } \{t_i' + \varphi_i; t_{i+1}; t_i' + T_i + e_i\}$; for $i = 1, 2, 3\ldots$

The recursion terminates when $t_{i+1}' = t_{i+1} > 0$ because all phases encountered from then on are of normal duration.

References

Aimsun, 2018. Aimsun Next Version 8.3 User's Manual. Barcelona, Spain www.aimsun.com.

Allsop, R., 1968. Selection of Offsets to Minimize Delay to Traffic in a Network Controlled by Fixed-Time Signals. Transportation Science 2 (1), 1–13.

Boeing, G., 2017. OSMnx: new Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks. Comput Environ Urban Syst 65, 126–139. https://doi.org/10.1016/j.compenvurbsys.2017.05.004.

Caltrans, 2020. Traffic Signal Operations Manual. [ebook] pp.65-67. Available at: https://dot.ca.gov/-/media/dot-media/programs/traffic-operations/documents/mobility/traffic-signal-operatons-manual-1-31-2020-a11y.pdf.

Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A., 1996. A modified logit route choice model overcoming path overlapping problems. In: Proceedings of the 13th international symposium on transportation and traffic flow theory. Pergamon, Elsevier Science Ltd, Oxford, UK.

Cascetta, E., 2001. Transportation Systems Engineering. Dordrecht: Kluwer.

Daganzo, C., Lehe, L., 2016. Traffic flow on signalized streets. Transportation Research Part B: Methodological 90, 56–69.

Daganzo, C., Lehe, L., Argote-Cabanero, J., 2017. Adaptive offsets for signalized streets. Transportation Research Procedia 23, 612–623.

Federal Highway Administration, 2005. Signal Timing on a Shoestring. U.S. Department of Transportation, Washington.

Gartner, N., Assman, S., Lasaga, F., Hou, D., 1991. A multi-band approach to arterial traffic signal optimization. Transportation Research Part B: Methodological 25 (1), 55–74.

Gartner, N.H., Little, J.D., 1973. The Generalized Combination Method for Area Traffic Control. Cambridge: Massachusetts Institute of Technology, Operations Research Center.

Gartner, N., Stamatiadis, C., 2002. Arterial-based control of traffic flow in urban grid networks. Math Comput Model 35 (5–6), 657–671.

Gartner, N., Stamatiadis, C., 2004. Progression Optimization Featuring Arterial- and Route-Based Priority Signal Networks. Journal of Intelligent Transportation Systems 8 (2), 77–86.

Gartner, N.H., Little, J.D.C., Gabbay, H., 1975. Optimization of Traffic Signal Settings by Mixed Integer Linear Programming. Part I: the Network Coordination Problem. Transportation Science 321–343.

Gordon, R., 1996. Traffic Control Systems Handbook. U.S. Dept. of Transportation, Federal Highway Administration, Washington, D.C.

Gregoire, J., Frazzoli E., de La Fortelle A., 2014. Back-pressure traffic signal control with unknown routing rates. Proceedings of the 19th World Congress The International Federation of Automatic Control Cape Town, South Africa.

Hillier, J.A., 1966. Appendix to Glasgow's Experiment in Area Traffic Control. Traffic Engineering and Control 7 (9), 569–571.

Hu, H., Liu, H., 2013. Arterial offset optimization using archived high-resolution traffic signal data. Transportation Research Part C: Emerging Technologies 37, 131–144.

Hu, H., Wu, X., Liu, H., 2013. Managing oversaturated signalized arterials: a maximum flow based approach. Transportation Research Part C: emerging Technologies. Volume 36, 196–211.

Little, J.D.C., Kelson, M.D., Gartner, N.H., 1981. MAXBAND: a Program for Setting Signals on Arterials and Triangular Networks. Transportation Research Record: Journal of the Transportation Research Board 40–46.

Little, J., 1966. The Synchronization of Traffic Signals by Mixed-Integer Linear Programming. Oper. Res. 14 (4), 568–594.

Liu, Peng., 2015. Development and Evaluation of a Network-wide Traffic Signal Coordination (NETSCA) Algorithm for Improved Operational Performance. United States, University of Akron.

Morgan, J., Little, J., 1964. Synchronizing Traffic Signals for Maximal Bandwidth. Oper. Res. 12 (6), 896–912.

Newell, G., 1981. Blocking Effects for Synchronized Signals. 8th International Symposium on Transportation & Traffic Theory, (pp. 469–489). Toronto.

Newell, G., 1989. Theory Of Highway Traffic Signals. Institute of Transportation Studies, University of California, Berkeley, Calif.

Pignataro, L.J., 1978. Traffic Control in Oversaturated Street Networks. National Cooperative Highway Research Program Report 194. Transportation Research Board, Washington, DC.

Quinn, D.J., 1992. A Review of Queue Management Strategies. Traffic Engineering and Control. November.
Rathi, A.K., 1988. A Control Scheme for High Traffic Density Sectors. Transp. Res. 22B (2), 81–101.
Robertson, D.I., 1969. Transyt: A traffic network study tool. RRL Report LR 253, Road Resarch. Laboratory, Crowthorne, Berkshire. https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=1173419.
Sadek, B., 2021. Synchronizing Network Traffic Signals on Grid Networks During a Rush. [Unpublished Doctoral Dissertation]. University of California, Berkeley.
Siksna, A., 1997. The Effects of Block Size and Form in North American and Australian City Centers. Urban Morphology, pp. 19–33.
Smith, M.J., 1979. The Existence, Uniqueness and Stability of Traffic Equilibria. Transport Research 13B, 295–304.
Traffic Research Corporation, 1966. SIGOP: Traffic Signal Optimization Program : a computer program to calculate optimum coordination in a Grid network of synchronized traffic signals. Traffic Research Corp.
Trafficware, 2012. Synchro Studio 8 User's Guide. Trafficware Inc, Albany, CA.
Varaiya, P., 2013. Max pressure control of a network of signalized intersections. Transportation Research Part C: emerging Technologies. Volume 36, 177–195.
Yang, X., Cheng, Y., Chang, G., 2015. A multi-path progression model for synchronization of arterial traffic signals. Transportation Research Part C: Emerging Technologies 53, 93–111.
Ye, B., Wu, W., Mao, W., 2015. A Method for Signal Coordination in Large-Scale Urban Road Networks. Math. Probl. Eng. 1–15, 2015.