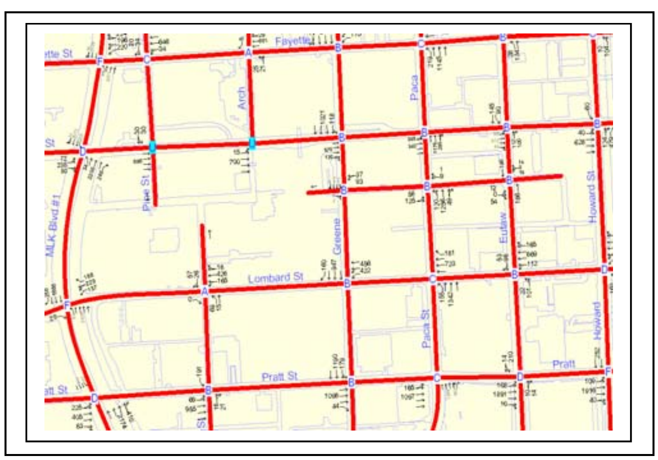
Maximizing Benefits of Signal Timing Optimization

Ziad A. Sabra, Ph.D., P.E., PTOE, Principal Traffic Engineer, Sabra, Wang & Associates
Keith Riniker, P.E., PTOE, Senior Traffic Engineer, Sabra, Wang & Associates

Abstract

In recent years, it seems that signal timing has become a tool-dependent and a less knowledge-based exercise. Based on the authors' experience, many users expect signal timing tools to produce readily used results and some use it as such. It is quite frequent that many traffic engineers who develop signal timing plans don't how controllers function, because they think it is someone else's job! This approach has many flaws in it, primarily because many practitioners who develop signal timing plans don't have the knowledge in signal controllers and may not know how to interpret the model's output and translate them into signal timing parameters in a signal controller. Relying entirely on models will not help users understand how signal controller-based improvements can be performed by utilizing functions inherent in controllers. Equally important is the knowledge and judgment that should be exercised when selecting cycle lengths, splits and offsets; these shouldn't be based entirely on what the models predict.

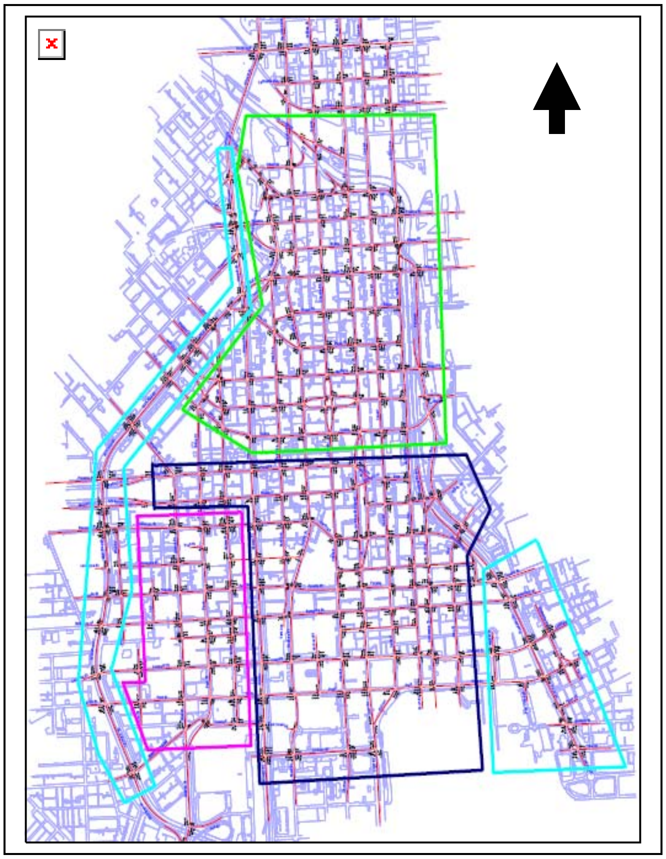
One would expect that the concerns above would be discussed in a Signal Timing Manual, but apparently that hasn't been the case yet. However, this knowledge can be obtained from peer-to-peer discussions, publications, and Signal Timing Forums. In fact, some universities (University of Idaho and Purdue University, for example) offer laboratory hands-on signal controller programming workshops, and ITE is also spearheading similar Webinars to spread this knowledge to young traffic engineers and signal timing technicians. This paper draws heavily on the experience used in a signal timing optimization project in Baltimore, Maryland consisting of 410 intersections and including a 275-signal CBD network, and others in Maryland and Virginia. Although the Synchro model was used extensively, significant improvements beyond the model's results were realized by using sound judgment in selecting number of timing plans; cycle lengths, splits and offsets; network partitioning; and in-depth knowledge of signal controller functions such as offset transitioning options; force-off options; dynamic split selection; left-turn conditional service; zero offset options; Transit Signal Priority (TSP), etc. This paper outlines and discusses practices that have proved to be significant in maximizing the benefits of signal timing optimization.



SIGNAL TIMING OPTIMIZATION AND EVALUATION TOOLS

A major question in signal timing optimization starts with: What signal timing model(s) should I use to optimize signals? A lot of attention should be given to selecting a signal optimization model. Users should address this question from several perspectives: credibility, simplicity, flexibility, and field friendliness. Whatever model we use should be credible and tested. It should be accepted by traffic engineers based on its efficiency and reliability of results. It should be simple to use by someone who understands the fundamentals of traffic engineering, and specifically signal timing but not necessarily a computer wizard. A signal optimization model should have built-in flexibility with user-defined adjustment/calibration factors to allow users to address driver behaviors and roadway geometric conditions. Lastly, the most appropriate model should be field-friendly with its inherent functions such as interactive time-space diagram and traffic flow dispersion profile; capacity analysis and level of service; and animated results. The most productive time of the signal optimization exercise, in my opinion, is in the field when putting the final signal timing fine-tuning touches to make things work just right! A model's animation and interactive capability combined with the user's signal timing know-how knowledge makes the signal timing optimization process very interesting and challenging. In our project, although several models would have met the criteria above, we selected the model Synchro based on its efficiency and field-friendliness; our traffic engineers were most comfortable with it and have had prior experience using it for field fine-tuning. On average, each intersection required approximately three hours to code and verify all data entry for three peak periods.

A major strength and advantage of Synchro over other available optimization tools, specifically to our project, lies in its ability to analyze whether a signal network should and can be partitioned into multiple optimization zones; secondly, in the flexibility of its data entry file structure which allowed multiple users to perform the data entry function simultaneously. The fact that several people could code the signal network (link-nodes and lane configuration, traffic volumes, phasing and timing) simultaneously, and subsequently combine all data into a single file, which for a 410-signal network was very important to the project schedule. However, it required a rigorous quality control procedure by a principal traffic engineer to make sure the multiple data entry correlated properly.



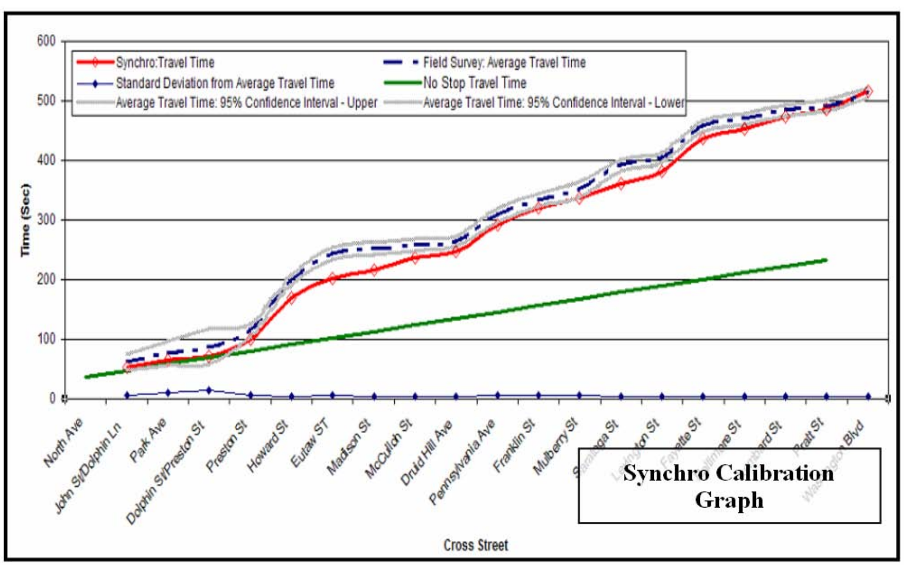
NETWORK ANALYSIS, PARTITIONING AND OPTIMIZATION

For a long time, signal optimization tools with logic to analyze multiple zones within a large signal netwrok did not exist, not even analyzing if some signals within a network should operate in an isolated signal timing "free" mode. Users most often used, and still use, the same cycle length for an entire network regardless of the implications, and wouldn't necessarily know what the implications are unless options are presented analytically. In many cases, the implications are unnecassary stops and delays, excessive fuel emissions, potential for crashes, and drivers frustration which often leads to questioning the credibility of traffic engineers and signal timing practitioners. Dealing with signal networks in large central business districts is a huge undertaking, and the use for anlytical tools to assist in making the right decision should not be compromised. Our judgment and intellactual ability have limits when it comes to these undertakings, and we must rely on tools to help us analyze practical options. In Baltimore, Maryland, the entire CBD with a 275-signal network used to run under a single cycle length during each peak period, regardless of the geometry and number of phases at the intersection. In those years, tools to analyze alternative options did not exist, and therefore, the most practical cycle length based on a model's optimization logic was used. Under our project, we used Synchro's^{TM} **"Partition Network"** function. Synchro's logic for subdividing a network is based on minimizing delays and stops as a result of operating multiple efficent cycle length for various zones. We must keep in mind that the model is only a tool, and as such a user could end up with numerous zones in a network. Therefore, sound judgment and knowledge of the network traffic patterns and roadway configuration is required to manually combine multiple small zones with cycle lengths that satisfy the requirements of all small zones. Under our project, we initially started with 30 zones in the CBD network. However, several zones were logically combined without adding delays and stops in the network. The final analysis and optimization of the network partioning resulted in five zones as shown in the figure above: 1) the MLK / Howard Street corridor, 2) the President Street corridor, 3) the CBD North, 4) the CBD South, and 5) the CBD West. This partitioning effort resulted in approximately 15% reduction in delays, measured in Synchro between the "single cycle" versus "multiple" cycle length analysis and optimization. The optimum cycle lengths varied from 80 to 110 seconds in the peak hours and 60 to 90 seconds during the off peak hours.

In summary, the benefits of partitioning the study area into multiple zones were to reduce cycle lengths, provide smaller areas that enable easier signal timing fine-tuning and routine maintenance, provide for easier updates of signal timings, reduce stops and delays, and improve overall signal coordination.

SYNCHRO CALIBRATION

The signal timing model calibration process is an important element of making sure that the base model, developed based on the existing signal timing plans, is accurate and realistic. Calibration of signal timing optimization models such as Synchro will take into account the variability of saturation flow rates as they may be affected by the intersections' characteristics, composition of traffic volumes, grades and



Synchro Calibration Graph

lane width, turning speed, etc. The figure here, for example, demonstrates how the model's derived travel time for one of the arterials was compared to several field-measured travel time runs, with a statistical confidence interval of 95%. Establishing a statistically valid model as part of the calibration and validation process is very critical before moving forward into the development of alternative signal timing optimization plans. After adjusting the parameters based on field-measured data and observations, our Synchro model was utilized to generate network and arterial performance measures; the Synchro results were further simulated in SimTraffic to visualize and validate existing conditions. The measures of effectiveness used to compare the field-measured conditions with Synchro/SimTraffic included total travel time, stopped delay, and queues for each arterial.

CYCLE LENGTH SELECTION

Some practitioners argue that long cycle lengths increase delays, stops and queues, while others argue that short cycle lengths cause more drivers to enter intersections on red, reduce length of progression bands, and penalize the major traffic movements. So, how do we quantify long and short cycle lengths, and what is the right cycle length? After all, the majority of cycle lengths in use at signalized arterials and networks are the result of the various signal timing optimization models.

The difference in opinion on the ***Cycle Length*** issue is more related to the philosophy of signal timing rather than on short versus long. Those who truly believe that maximizing coordination on major thoroughfares should dictate the determination of the optimum cycle length usually favor long cycles. Under these circumstances, the coordinated traffic movements receive more green time than the side street movements, and those who are delayed on the side streets usually will make up the delay if they are destined for the coordinated movements. Also, the ratio of the sum of fixed time intervals (Y+AR) to a cycle length is lower, when a cycle length is increased, than the same ratio for a shorter cycle length. Therefore, one could argue that increasing the cycle length for progression is more efficient than reducing it, and also has a lesser occurrence probability on the type of crashes that are associated with the change intervals. Very often, a progression-based cycle length is in the range of 1.3 to 1.5 times the model-determined optimum cycle length. This is one of the reasons why a typical progression-based cycle length for a linear arterial is not very sensitive to changes in the traffic peaking characteristics. However, in typical closely spaced signals on an arterial or in a grid network, a cycle length that is 1.3 to 1.5 times the model's optimized cycle length could result in congestion and blockage to upstream movements, unnecessary delays and stops, and a penalty to all road users.

At a minimum, except for coordinated signal systems, a cycle length should be the summation of all minimum timing intervals required to accommodate road users such as passenger cars, transit and pedestrians. For coordinated signals, the minimum cycle length should include any other time addition required for the offset seeking method used. Otherwise, phases will be omitted from the offset seeking transition and could result in sluggish and longer transition periods. Any duration that interferes with the absolute minimum phase times could result in the signal being out of coordination to a free mode. In some cases, an intersection with multiple signal phases may require a minimum of a 90 to 100-second cycle. ***So, would this be long or short?***

As practitioners we often generalize our opinion on the length of the cycle length. Signal networks are uniquely different among jurisdictions and their signal operations are driven by many constraints such as: storage length of turning bays, number of lanes, non uniform surges in traffic demand during the peak hours, pedestrian movements, safety, width of intersections, adjacent land uses, speed, number of signal phases, etc. Needless to say, local and state policies also dictate some of the signal operational requirements such as minimum green times and minimum pedestrian intervals. The right cycle length, regardless of its length, should serves all users safely and efficiently, and take into consideration the interaction of vehicular and pedestrian movements. A cycle length also should minimize the interaction between unserved vehicles, if any, with other vehicular and pedestrian movements, such as the case in a spill over. The right cycle length should have a reasonable expectation of failing, i.e. not able to serve all movements under unusual circumstances, but it should accommodate frequent surges of unusual peak hour traffic without sacrificing safety. Modern signal controllers, NEMA and/or Types 170 and 2070, have several functions to enhance optimized signal timing. Detection, being the core engine, in combination with controller functions such as conditional service for left turn, conditional re-service for through movements, adaptive split control, simultaneous gap out inhibit, traffic responsive selection, and other functions can add a great value to the management of the cycle length. The big question is whether most people who develop signal timing plans in the office know about these functions, understand how they operate, and be able to program the controller correctly with their selection and timing. Unfortunately, most often the answer is less than desirable!

Under our project, the City of Baltimore's main objective was to serve all users equally and to focus on the network total stops and delays throughout the CBD. Another goal was to hopefully select a cycle length that is shorter than what they had been using for many years, which varied from 110 seconds to 165 seconds during the peak periods. The problems resulted from the existing cycle lengths included queuing, several cycle failures during the peak hours, long cycle failure recovery time, and excessive delays to pedestrians who cross the major streets. Therefore, to meet these goals our priorities for selecting the ***right*** cycle length were clear: reduce queuing problems associated with the 400-foot spacing of signals on the grid network, select a cycle length that can accommodate fluctuation in traffic demand, and reduce pedestrian delays. Progression was not necessarily our top priority. Rather than selecting the Model's optimum cycle length, we actually used Synchro to identify and analyze a range of cycle lengths for each zone in the CBD. The Synchro's™ "**Optimize Network Cycle Length"** tool was used to compare queues, delays, stops and level of service under several different cycle length scenarios, and subsequently used the simulation engine in SimTraffic to evaluate the microscopic interaction among vehicles in the network. One of the early findings of our analysis was that a side street split would often be higher than necessary to accommodate pedestrians crossing the mainline roadway. A competing requirement was to provide adequate green intervals to process traffic on the mainline approaches while also reducing the cycle length. Due to a competing nature of these requirements, the minimum practical cycle length in the CBD during the peak periods was 80 seconds.

OFFSET SEEKING TRANSITION LOGIC

Offset seeking is commonly referred to the process used to transition from one signal pattern to another. When changing between signal timing patterns with different cycles and / or offsets, there are several different methods of transitioning with modern controllers. In Baltimore, using the Naztec Controller, the City uses the “Short/Long” transition method, and excludes the coordinated phase from shortening. The “Short/Long” transition method is an internal logic to the Naztec™ controller that calculates the shortest time-path between signal timing patterns. “Short” refers to the logic where the phases/splits are reduced between patterns, and “Long” refers to the logic where the phases/splits are extended between patterns. While the controller is in transition, coordination may not be guaranteed. Therefore, the quickest method of transitioning is normally preferred. With the “Short/Long” transition method, there are user-defined inputs for how much the phase duration can vary from cycle-to-cycle. These inputs are measured in percentage of cycle length. The Offset Seeking Transiting Logic effects the split and cycle length calculations. Each split needed to be 2 to 3 seconds greater than the minimum split to provide for transitioning between signal patterns. The split must be higher than the absolute minimum in cases where the controller selects the “Short” method and the actual split is reduced. Since the minimum cycle length is a function of the minimum splits on each phase, this setting could affect the minimum cycle length that is actually implemented. Understanding the mechanics of signal controllers is critical to the operation of a coordinated signal systems. The lack of this understanding will often render signals ineffective during coordination, i.e. some will be running in a **“free”** mode and thus compromising coordination and efficiency, and most likely safety. Young practitioners who haven’t had their first hands-on experience with programming controllers and field implementation should be aware of the controller logic for the offset seeking transition method.

SPLIT OPTIMIZATION

In most cases, the models’ derived splits are usually long enough to service the uniform arrival of traffic demand during the plan period, and accommodate a desired two-way progression. Selecting the right splits, however, requires good familiarity with all critical movement in the network. A critical movement is defined as one that can affect several other related movements in a negative way. For example, a short split for a left-turn movement with a less than desired storage bay could result in a spillover that would effect the adjacent and upstream through movements, thus causing unusual lane utilization at and upstream of the intersection. Such an occurrence could compromise efficiency and safety for these movements. Therefore, at a minimum, the user should know the maximum queues that critical movements can tolerate, and the implications of a cycle failure. There are times when queue management for critical movements is far more important than attaining a theoretical progression.

For critical movements that experience unusual surges of traffic during a portion of the peak hour (e.g., school entrance/exit, parking garage, theaters, restaurants, shopping centers, etc.) it is best to optimize its split based on the peak 15-minute arrival. For example, a movement with 100 vehicles per hour, which consists of a one 50-vehicle per 15 minutes should be entered and optimized in the model based on 200 vehicles per hour. This requires that the critical movement have proper presence detection.

As a start, a user should determine the minimum green time that each movement at the intersection requires (i.e., greater of the G+Y+AR and WK+FDW+Y+AR). Additionally, the capacity/maximum number of actual queued vehicles should be determined as well. The minimum spilt should be entered in the signal optimization model as a minimum requirement. The maximum queue should be used to compare with the Model's output. Once the model's splits are optimized, the user should check each optimized split separately and determine the maximum queue associated with it. If the maximum queue is longer than what the movement can accommodate then the split should be adjusted upward. A starting point for the desired split time can be performed manually (Volume per Lane /Number of Cycles per Hour x Surge Factor of 1.3 to 1.5).

In some other cases, for instance when using Synchro, the optimized splits may be higher than required when traffic demand for a movement is at or near capacity. Synchro uses the percentile method to derive an average split time based on five percentile occurrences, including when the demand is several standard deviations above the mean arrival rate, i.e. in the 70 and 90-percentile conditions. Therefore, in cases where demand equals to capacity, the 50-percentile arrival condition will occur very frequently, the 70-percentile arrival condition may occur very infrequently, and the 90-percentile traffic condition has zero probability of occurring. Under these circumstances, the user should be aware of these conditions and should analyze the splits carefully. Simulation of the results is always desired to evaluate timing for saturated conditions. Macroscopic models such as Synchro, Passer and Transyt are better suited for optimizing traffic conditions that are at or below capacity. However, they can be as effective if the optimized signal timing plans are enhanced further based on the results micro-simulation models such as SimTraffic, CORSIM and/or VISSIM.

In summary, split optimization may require another visit to the cycle length optimization so that all signal timing settings are harmonic.

OFFSET OPTIMIZATION

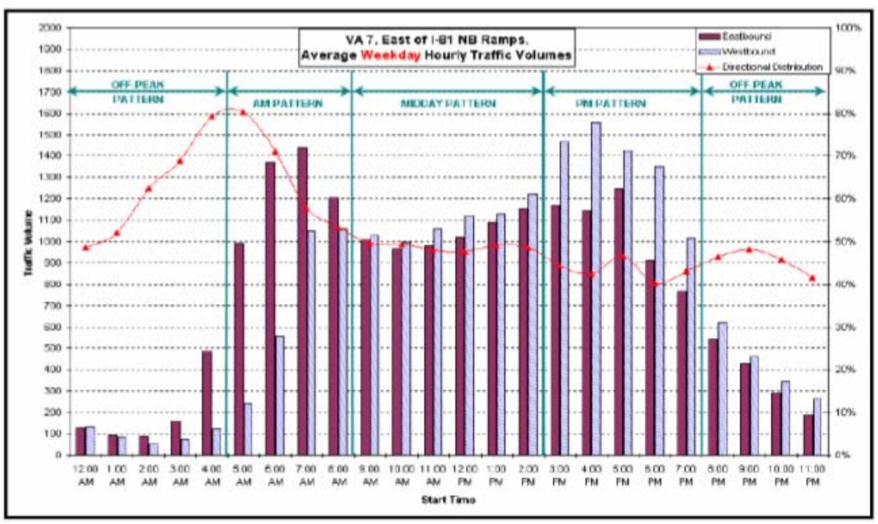
Offset optimization is by far the most challenging to determine of all signal timing settings. Offsets effect progression between signals, and have no effect on the theoretical intersection level of service and capacity. Offset, however, have a major effect on congestion management and throughput of a system. Cycle length and splits tend to have some tolerance built within their ranges, specifically to manage the increase and fluctuation in traffic demand. Offsets, however, are based on a pre-determined directional split of traffic and usually do not have much tolerance to accommodate more than a 10-percent change in the directional split. Also, unlike splits, offsets associated with fixed time-of-day signal plans will not change regardless of the traffic demand conditions unless the signal system is operating in a traffic responsive mode. For example, a 70/30 directional split in the favored northbound direction will not produce the same throughput as in a 60/40 split condition. Multiple offset plans, with the same splits and cycle lengths, could be used by time-of-day to address changes in traffic patterns.

Offsets should be optimized based on a logical priority basis. Although signal timing optimization models optimize offsets for an entire network in a single run, the user still has the option to optimize offsets for arterials on a selective basis. For example, in a network of 10 arterials, one can establish the offsets for the top five arterials by setting priorities from 1 to 5, with 1 being the most critical arterial and the last five arterials being equal. The most critical arterial(s) can be optimized first and the selected offsets can be fixed in the model while optimizing the offsets for the second arterial. This process can be repeated in the model until all offset priorities are met. This process was followed in the signal optimization project in Baltimore and produced favorable results.

One important thing that users should be aware of is the utility of the optimization and simulation tools. These tools provide us the processing power we never had before, to test multiple scenarios very efficiently.

NUMBER OF SIGNAL TIMING PLANS

In general, the number of signal timing plans should be as minimum as particle. The greater the number of plans the more maintenance would eventually be required to upkeep them. Nevertheless, the number of timing plans should correspond to the significant variation, i.e. lows and highs, in traffic volumes during a typical week, including weekend days. Often, agencies may develop one or two additional timing plans for inbound and outbound special events such as for parades, emergency evacuation, conventions, sport events, etc. The figure below demonstrates a typical fluctuation of traffic demand during a typical week and the average durations that can be covered in a single timing plan. It is frequent that the off-peak signal timing plans plan could be used for all hours that are not covered in the morning and evening peak plans, i.e. midday, nights, early morning, and weekends. A morning peak signal timing plan usually handles traffic demand between the hours of 6:30 am and 9:30 am. Likewise, an evening peak signal timing plan usually handles traffic demand from 3:30 pm to 7:30 pm. The off peak plan usually handles all other hours. Weekend plans could use same weekday morning and evening peak plan but with different time duration and different offsets to accommodate the variation in the direction split of traffic volumes. In the City of Baltimore, three timing plans were developed and are in use today for the typical week operations. However, the City also developed a few other special event timing plans for the various seasonal events that take place in and around the City. Users should be aware that signal controllers take about 3 cycles to transition between signal timing plans while the signal may be out of coordination; therefore, plan transition should avoid any part of the peak period if all possible. It is usually better to combine the two plans into an average of the two plans to avoid transitioning. This strategy is evidently successful since most practitioners use a three or a four-hour peak period for the AM and PM signal plans.



Sample Traffic Volume Analysis Graph

TRAFFIC RESPONSIVE PLANS (TRSP)

Traffic responsive plans are geared to signal systems that experience frequent non-recurring fluctuation in traffic demand, or any other condition that is not considered predictable by time-of-day. Examples of those conditions include incidents, early release of sporting events, sporadic arrivals and departures at special events, evacuation of beach traffic, detours, emergency evacuation, etc. The advantage of implementing TRSPs versus implementing multiple time-of-day plans lies in the concept of "Signal Timing Plan Responsiveness", as may be triggered by system detector data (traffic volume and occupancy) caused by the variation in traffic demand.

Typically, a signal timing plan (Cycle, splits and offset) stays in effect until a new threshold of volume plus occupancy (V+KO) is reached for a reasonable period of time. A new signal timing plan is then activated with either a higher or a lower cycle length and splits, and same or different offsets as well (based on directional distribution of the traffic volumes). This process can be in effect all day long to serve the traffic demand more responsively. The greatest advantage of operating such a strategy is in its ability to implement flexible signal timing plans based on a ***"near"*** real-time traffic demand rather than a preset time-of-day timing plan. A preset time-of-day signal timing plan does not have the ability to respond to an unexpected surge or sink of traffic volumes, as may be induced by incidents, lane closures, emergency evacuation and traffic conditions that may change unexpectedly due to inclement weather conditions. Nevertheless, a TRSP plan is always better than a fixed time-of-day plan, but may not always be responsive unless a significant amount of time is spent on the sensitivity analysis of setting the V+KO thresholds. Choosing insensitive thresholds of V+KO can cause the signal controller to stay in transition for a long time, thus negating the benefits of implementing several flexible signal timing plans. On average, a signal controller can require up to three cycles to get back in step with coordination when transitioning between signal timing plans; only applies if the offsets are non zero. For example, using a cycle lengths of 120 seconds, a controller may require 6 minutes in transition (running without real coordination) before it synchs in coordination. Using a short system detector count period, coupled with insensitive TRSP thresholds, could compound the complexity of this strategy. On the other hand, using sound judgment with realistic outcomes, the TRSP could be applied very efficiently to serve the purpose of selecting flexibility in responsive signal timing plans.

The determination for effective TRSP can be tested using the "KREER" approach, dated back to the 1970s, which states that if the Coefficient of Variations (CV) of traffic volumes is greater than 20%, then TRSP will produce reduction in stops and delays by at least 5% better than an optimal time-of-day timing plan. This theory was used generously with the old UTCS system and functioned very well for many agencies, and is still advocated even now. The CV is determined by the Standard Deviation of the traffic volumes divided by the Mean. The ***Standard Deviation*** is a measure of the volume variability for the same times of day on multiple days. The ***Mean*** is the average of the volumes for the same times of day on multiple days.

Traffic responsive plans are not as difficult to develop and implement as it may sound. Many agencies, for example Maryland, use TRSP plans generously and have attained excellent results; they have been using TRSP plans for more than 50 arterials for more than 10 years. TRSP plans are developed just like the time-of-day plans, except they encompass more timing plans to accommodate higher and lower thresholds of traffic demand, and higher and lower directional distributions of traffic. A TRSP plan matrix may include 10 plans to choose from. TRSP plans can be tested with simulation models to determine their effectiveness before they are implemented in the field.

FIXED VERSUS FLOATING FORCE-OFF

Force offs determine the termination points of all programmed phases. The selection of the force-off method in the controller unit (fixed vs. floating) can make a major difference during coordination, and could impact driver expectations and perception of signal coordination. With fixed force-offs, non coordinated phases, say phases 4 and 8 for the side streets, would extend their green time, i.e. start early and use all remaining time of the preceding gapped out phases, say 3 and 7. This phenomenon occurs because phases 4 and 8 must comply with the fixed force-off points for those phases. Under this setting, the side street and not the coordinated phases benefits from the additional green time, but also could unload vehicle to the downstream intersections which may or may not be a desired situation. The advantage of the Fixed Force-off method, however, is that the coordinated phases will always start and end at the same point of the cycle, regardless of what happens on the side street (unless there is no demand); drivers expect this routine in signal coordination and give them a sense of good coordination.

A Floating Force-off selection in the controller works exactly the opposite of the Fixed Force-off selection. Any time that remains as a result of a gap out in the non-coordinated phases will return to the coordinated phases, and each non-coordinated phase will only receive its max time unless it gaps out. For example, if phases 3 and 7 gap out early and there is 10 seconds left in their max timer, then phases 4 and 8 would be serviced next in order for the full duration of their maximum time only unless they gap out early as well. Subsequently, any collective time that is not used by phases 3, 4, 7, and 8 would be allocated to the coordinated phases 2 and 6. Therefore, phases 2 and 6 would receive an early release of the Green display for that amount at the subject intersection. The disadvantage of the Floating Force-off is that the ***"early release"*** of green to the coordinated phases may cause a problem at the downstream and upstream intersections, i.e. an additional queue build-up that may cause a shock-wave effect to the coordinated movements; also, the early release may fail driver's expectations when they stop at the next intersection thus raising a doubt as to whether the signal systems is properly coordinated.

So, which method is best? There is no affirmative answer to this question. The users should clearly understand both methods and weigh their application on an intersection-by-intersection basis, based on the characteristics of the intersections and traffic demand.

RESERVING A LEFT TURN TWICE DURING THE CYCLE (CONDITIONAL SERVICE)

There are conditions when a left-turn movement may need to be serviced twice during the cycle. This controller function is very useful for left turns that experience heavy queues, short storage bays and also unusual arrival patterns. Conditional service, when selected in the controller unit, may allow a dual ring controller to re-service the odd phase after the even phase but before crossing the barrier. This function is typically used with all 2-ring, 2-barrier type controllers, using all odd phase 1, 3, 5, 7. Conditional service requires presence detection for the left-turn movement and advance detection for the main street movements. Conditional service is usually allowed as long as three conditions are met: ***1)*** the even phase in the same ring has gapped out and is resting; ***2)*** there is a call across the barrier (without a call Conditional Service is not meaningful); and ***3)*** the even phase in the opposite ring is still extending, and there is enough time left in its maximum timer that is greater than the "conditional service minimum time" plus the even phase's, same ring change interval time.

CONDITIONAL PED SERVICE

The same concept of conditional service above also applies to pedestrians. For example, if there is no call for pedestrian service at the beginning of the phase, the timer will allow late pedestrian calls and will re-service the pedestrian displays (WK+FDW) as long as there is enough time in the maximum timer for the compatible phase. This function is very useful to service pedestrians responsively and minimize jay walking.

CONDITIONAL RESERVICE (LEFT TURN, NEMA TS1)

The State of Maryland, being one of the leaders and innovators in signal timing management, also uses, at several intersections, the left-turn re-service function with multi-ring, multi-barrier NEMA TS1 controllers to manage left-turn demand and queues at intersections with constrained storage bays. This application doesn't use the three conditions listed above, and has helped reduce stopped delay and left-turn red signal violations.

CONDITIONAL RESERVICE (THROUGH MOVEMENT)

The Conditional Re-service is another function that would allow the timer to return to the through phase after the left-turn Conditional Service is serviced, otherwise the timer will continue to time the left-turn phase then clear to the serve phases across the barrier. Under this condition, the Maximum times apply.

SIMULTAENOUS GAP-OUT INHIBIT

Under normal cases, all phases on one side of the barrier must wait for one another before crossing the barrier, i.e. they must clear the phases together. For example, if phase 6 gaps out but phase 2 is still extending, then phase 6 will rest in green and wait on phase 2 to gap out as well before both phases begin to time the change intervals. However, if phase 2 gets another call while it is resting and before the phase max time is maxed out, then the timer will reset and start timing again. This process, for non-coordinated signals, can repeat itself over and over. Such an occurrence can result in a sluggish operation at an intersection. The Simultaneous Gap-Out Inhibit function, if selected, will prevent the timer from extending a phase after it has gapped out, thus resulting in a snappier operation since the barrier is crossed as soon as the other compatible phase gaps out as well. Note this controller function requires detection for all even phases associated with the Simultaneous Gap-Out Inhibit function.

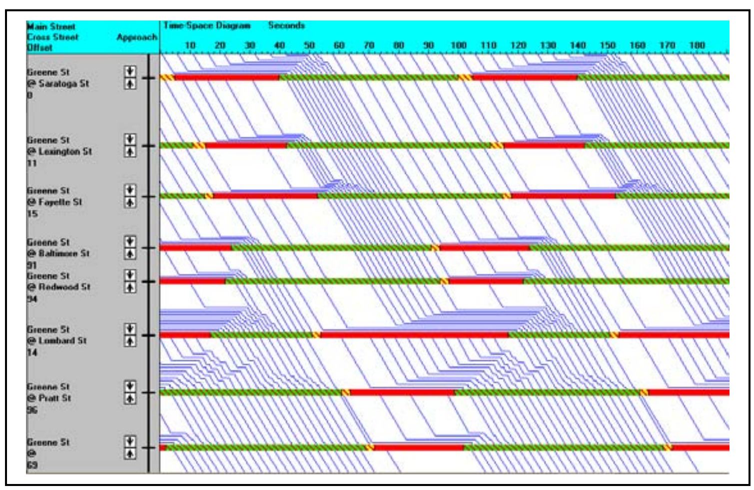
LOCAL DETECTION

Not enough can be said about the need for good vehicular and pedestrian detection. The absence of presence detection at signalized intersections compromises the integrity of signal timing optimization. *The myth of "intersection operation with detection during peak hours is equivalent to operating with a fixed-timing" is not true.* Signalized intersections don't operate only during peak hours, and they don't always operate at capacity. They operate 24 hours a day, 365 days a year. In my opinion, nothing can substitute for presence detection at signalized intersections. The efficiency of signal timing at a signalized intersection can be diminished by more than 30-percent without local detection. This can be easily tested in any signal timing model under various traffic demand conditions. Without a proper pedestrian detection, i.e. push buttons to cross major streets, the pedestrian movement would have to be placed on recall for its full duration; in many cases, it results in timing a phase without any visible demand of vehicles and pedestrians. Likewise, a left-turn or a side street through phase would have to time a full duration when detectors malfunction or are absent. Presence detection at signalized intersection is very critical to the intersection operation and safety, and more so for drivers' trust in our ability to time signals properly. Therefore, before beginning a signal timing optimization project, make sure that intersection presence detection is addressed adequately.

When all fails and you are still searching for a few seconds to increase capacity for the mainline movements, it isn't too late to examine in the field the effect of shortening the gap extension time for the side street movements. Shortening the gap extension time for side street movements, for speeds below 35 mph, can increase efficiency and throughput. However, factors such as sight distance, grades, vehicle composition, rear-end crash history, and overall safety at the intersection should be studied as well to determine the trade-offs.

TIME-SPACE DIAGRAMS (TSDs)

A TSD is a very fundamental principle in signal timing. The abundance of computerized interactive time-space diagram tools, however, has made the concept of TSDs more difficult to understand by young engineers, compared to the days when this exercise was done manually and was well understood. Most of the thinking is now replaced by the power of computers. Nevertheless, interactive TSDs are very useful for signal timing fine-tuning and preventive maintenance. Many small jurisdictions use TSDs to optimize offsets for coordinated signal systems. TSDs are very productive in the field for the initial fine-tuning effort. It allows us to compare the model-derived progression bands and those that exist in the field, either based on existing conditions or by programming of the new offset values. TSDs are very useful for verification of offset values in the field, and can easily pinpoint an erroneous data entry of offsets and phase sequences. Knowing how to program the controller with new timing plans would allow the user to make field adjustments in the streets and see immediate improvements. TSDs should be used in the field to verify and adjust the operating speeds versus the coded speeds in the model. Speeds should be verified on a link/block-by-block basis and should be adjusted in the model accordingly to reflect the as-is conditions. Progression is very sensitive to changes in offsets. For example, a change from 30 to 40 mph over 1,000 feet will change the travel time by 6 seconds. Compound this mistake over a mile and the offsets would be off by 30 seconds, which can result in a significant effect on progression and stopping.



TRANSIT SIGNAL PRIORITY (TSP)

Under normal operating conditions, the TSP functions well, as long as there is enough slack time in the cycle length, specifically for the transit conflicting movement, that can be utilized for the TSP function. Under unusual conditions, phases may be skipped or shortened and, therefore, more time is usually available to accommodate TSP calls. TSP, being a priority function, works under the principle that the controller timer will allocate unused time from the transit conflicting phase to the transit parent and/or compatible phase, yet keeping the timer in coordination. For example, if a transit vehicle places a call at the intersection while the parent phase is active but still needs 10 additional seconds, then the timer will extend the TSP parent phase and subsequently either skip one of the next phases or shorten the phase by 10 seconds. Some agencies don't allow skipping a left-turn phase or even shortening the Walk interval. Most agencies would accept serving only the minimum duration for a phase, and /or skipping the pedestrian WK and FDW display in order to allocate time for TSP. In Baltimore, additional time was built into the cycle length and splits on side streets to allow this slack time to be utilized for TSP. This approach has worked well so far and is a good compromise to move all traffic equitably.

LOCAL DYNAMIC SPLIT CONTROL

The term dynamic spilt control is used sometimes interchangeably by controller vendors with adaptive split control. This feature, which exists in almost all NEMA and Model 2070 controllers, allows each intersection to select its own split based on actual vehicular demand. Although this function works almost the same in all controllers, there are still some application variances between vendors. In general, this is how it works. Each split is assigned a set of selective phases, and force-offs are monitored by the controller for each of the phases over a number of cycles. The user will identify several local dynamic splits, say splits 1, 2, 3 and 4. Each split would have an assigned time as well. When a pre-specified number of force-offs is reached then the controller will assign a pre-specified split timing to those phases that exceed the threshold value. For example, let us assume that dynamic split 1 has phases 3 and 7 assigned to it. Assume that phase 3 forced-off three times and phase 7 forced-off two times over a number of cycles. Therefore, the total number of force-offs would be 5. Starting with split 1, the split with the highest total exceeding the minimum threshold is selected for the duration of the next sampling period. The controller will remain in the current split until a new split with a higher number of force-offs exceeding the threshold is determined. If two splits have the same number of force-offs than the lower numbered split will be selected. In a coordinated signal system, the dynamic split always steals time from another phase in order to stay in coordination.

SETTING EXPECTIONS WITH COORDINATION AND OPTIMIZED SIGNAL TIMING PLANS

When developing signal timing plans, it would be reasonable to set some expectations at the beginning of the project. For example, below is a list of expectations that have proved successful in our projects:

- Establish goals and minimum expectations for improvements. Here is list of typical measures of effectiveness: stops, delays, air emissions, speed, number of driver complaints, queues, number of cycle failures, level of service, etc. Many of these metrics can be determined based on the preferences of the agency. For example, in Baltimore, the goal was to reduce total stops and delays and improve operations and safety for pedestrians.

- Always consider safety and risk analysis to all users at intersections, and the implications of innovative strategies for signal timing optimization.

- Know all unique lane utilization configurations for each intersection in the network. Not all lanes are created equal. At least not in the eyes of the driver. They prefer to use some lanes more than others. You need to know this observation when you are developing your signal timings. Maybe it's a double left turn and no one uses the outside lane! Lane utilization effects capacity and capacity effects split optimization significantly.

- Identify the minimum change intervals for all movements using the ITE prescribed method. Doing this activity at a later stage can cause a major "redo" of signal timing optimization.

- Identify the minimum WK+FDW for all pedestrian movements using the ITE prescribed method. Doing this activity at a later stage can cause a major "redo" of signal timing optimization.

- Understand the default settings in your controllers. Becoming familiar with the offset seeking methods, programming of force-off selection, and special functions, etc. is very critical to your ability to enhance the signal timing plans with functions that would serve some movements and intersections.

- Know if your agency would allow you to experiment with changing the left-turn phase sequence. Also, consider lead-lag for left turns and understand what constitutes a yellow-trap. The Yellow Trap doesn't exit with a protected left-turn (3-section signal head) phase. Consider changing lead-lag operations by time of day if it provides a larger green band.

- Plan your stopping points in a coordinated signal system to occur at some of the same intersections motorists used to stop before. Although, new signal timing plans may reduce delays and stops, however, drivers may continue to complain and assume that the new timing plans are not effective if they stop at intersections that they did not stop before. Stopping at the same locations, but for less time, is more acceptable to drivers. Stopping at low volume intersections is not desired either.

- Know all movements that have malfunctioning local detectors and attempt to correct these problems before the new signal timing plans are implemented.

- Identify the most critical intersection(s) in your coordinated system, typically the cross intersection of two coordinated arterials, and assign a "zero offset" to this intersection so that it doesn't get out of coordination when transitioning from one timing plan to another. In a typical NEMA and Type 170 controller, it is normal to expect a minimum of 2 to 3 cycles to transition from one timing plan to another. With a 2-minute cycle, this could be a total of 4 to 6 minutes when the controller is out of coordination.

- For choked progression bands, users should consider restricting shared left-turn movements, by time-of-day, if reasonable alternatives exit for the left-turn movement. Would your agency allow this mitigation strategy?

- Consider a delayed output selection for the left-turn movements if permissive/protected phasing is used. This will improve the efficiency of the intersection for low volume left-turn movements, especially when gaps in the opposing direction are plentiful. Some agencies even consider placing the stop-line detector 30 to 40 feet upstream from the stop line and only call the phase if queues extend to the detector; otherwise, the signal stays in the permissive display.

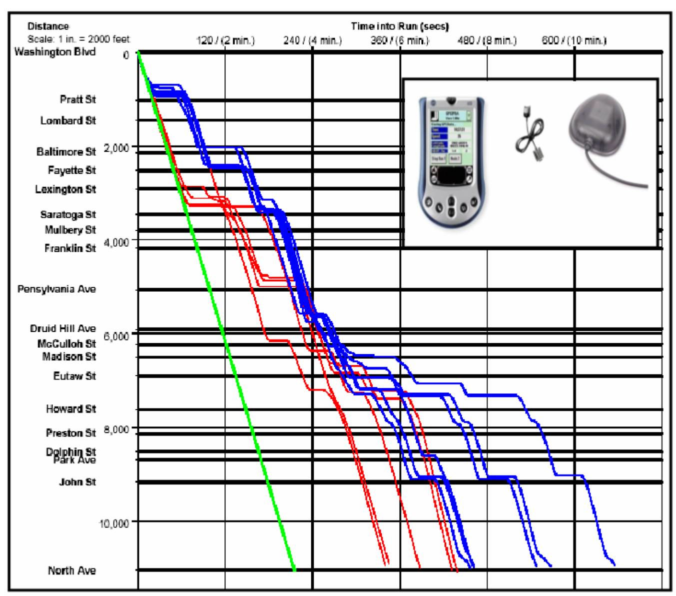
- When developing signal timings, spend a lot of time making sure you have the right cycle length. This is the most difficult parameter to fine tune in the field. Test the sensitivity of your timing plans and cycle length for a 5%, 10% and 15% change in traffic demand. The cycle length should be able to tolerate at least a 10% change in traffic demand. If not, then reconsider the cycle length selected for your timing plans.

- Expect to have slack time in the transit conflicting phases if TSP is implemented; therefore, a longer cycle length would be required with TSP. Without slack time, the TSP function will be very restricted and may not be effective.

- Signal timing models should be calibrated before using them for signal timing optimization. More on the calibration methods can be found in the FHWA Traffic Analysis Toolbox, Volumes 1, II, III, June 2004

- Focus more on queue management and movement failures than on delays and stops when dealing with saturated traffic conditions. The lack of addressing queue clearance and cycle failure can create major bottlenecks in the signal network, and sometimes can cause a ***"true gridlock"*** that would hinder movements at several intersections.

- Travel time delay is one of the best performance measures for coordinated signal systems. Drivers value their total travel time more than how much they stopped at a particular intersection. Expect to conduct



Sample Field-Collected GPS Data

travel time runs at least once a year for all major arterials in your network to determine potential degradation of signal timing plans. At least six non-impeded runs should be performed and averaged for each peak period, but it is better to follow a statistical approach to determine a statistical sample size based on a 95% level of confidence and a sampling error of 10% or less. A ten-percent degradation or higher in travel time is a reasonable measure to assume for retiming.

- Invest in a GPS-based travel time system for travel time studies; the most cost-effective tool a Traffic Engineer can have for signal timing studies. It only requires one driver and no manual activities in the car other than driving. Vehicle positioning data can be recorded automatically in $1/0^{th}$ second increments, then downloaded and translated into travel time, delay and queues within seconds. The total cost for a travel time GPS-based system including the receiver, PDA and software is approximately than $1,200. This investment can be recovered, labor time savings, in one or two studies.

Authors:

Ziad A. Sabra, Ph.D., P.E., PTOE is a Fellow Member of ITE. Ziad is a graduate of West Virginia University. He is a Principal Traffic Engineer at Sabra, Wang & Associates, Inc. in Baltimore, MD. Ziad has 26 years of experience in traffic engineering planning, design and operations. He is a registered professional Engineer in 10 states. He has led numerous signal timing optimization projects (over 2000 intersections) in his career, and also worked on several signal system design projects. He served as the Principal Designer and System Manager for the upgrade of the City of Baltimore 1300 signals system encompassing the replacement of controllers, communication hubs, the central system software and hardware and the Transportation Management Center. Ziad is very active in FHWA and TRB research projects, specifically in Light Rail Transit Signal Priority (TSP), adaptive signal timing strategies (ACS-Lite), signal timing performance measures, signal timing and traffic modeling, and traffic control equipment and computerized signal systems. Ziad is a recognized Peer with the FHWA Peer-to-Peer Program for Traffic Control Devices, and is a Principal Author of several research publications and training manuals for FHWA/NHI. Ziad can be reached at zsabra@sabra-wang.com, or by phone at (410) 737-6564, Ext. 103, and fax at (410) 737-1774.

Keith Riniker, P.E., PTOE is Member of ITE. He is a Senior Traffic Engineer at Sabra, Wang & Associates, Inc. in Baltimore, MD. He is a graduate of University of Maryland. Keith has 13 years of experience in traffic signal design, signal system design, signal system communications, traffic signal timing and optimization, traffic operations and safety studies and analyses, traffic data collection and analysis, traffic simulation and modeling, traffic control design, work zone traffic control, congestion management, and traffic management systems. Keith served as the Principal Traffic Engineer for the City of Baltimore Signal Timing Optimization Project. He is currently serving in a similar capacity for several other signal timing optimization projects in Maryland, Virginia and Pennsylvania. Keith is an expert user of several signal optimization models, namely Synchro, SimTraffic, CORSIM and VISSIM and has used these models for projects of various complexities including several miles of arterials and interchanges. Keith can be reached at kriniker@sabra-wang.com, or by phone at (410) 737-6564, Ext. 119, and Fax at (410) 737-1774.