MASTER THESIS

Optimization of coordinated traffic signal intersections

Author:
Thomas Borre Jørgensen

Supervisor:
Marco Chiarandini

Department of Mathematics and Computer Science
University of Southern Denmark, Odense

1. June 2018

Contents

Chapter 1

Background

1.1 Introduction and motivation

The purpose of traffic optimization is to address the task of improving the transportation of each entity in a traffic network from an origin to a destination, based on a set of optimization criteria. Overall the root of any issue has to do with traffic congestions due to demands exceeding the capacity of the traffic network. As of today this presents an increasing problem due to an also increasing number of vehicles, especially motorized ones.

Traffic signals at intersections is one of the most widespread means of controlling traffic. Very large cities may contain hundreds to thousands of intersections requiring traffic signals. Consequently, they are closely grouped together such that the improvement of a single signal may be at the expense of a large neighbourhood of its adjacent signals or even the entire traffic network of the city. Therefore, its is necessary to consider multiple signals simultaneously to allow coordination.

1.2 Terminology

We introduce the technical terms, notions and definitions necessary to the subject of this thesis.

A land based *traffic network* consists of roads on which traffic takes place. Any road consists of a number of traffic lanes. A *traffic lane* is a segment of the road having an appropriately width in order to fit only a single line of vehicles or bicycles, therefore it may serve traffic in a single specific direction only. The most common types of roads, in increasing order of capacity (more lanes and higher speed limits), are *streets*, *collector roads*, *arterial roads* and *free-ways*. A *traffic junction* is any point in the network at which at least two roads meet. Depending on the network topology the roads may meet at different levels of elevation. There are two levels of elevation that characterize the type of junction.

1. An *interchange* is a *grade separated* (over- and undergrade) junction. They involve roads passing below and over each other by means of tunnels and bridges. The exploitation of three-dimensional road networks gives rise to several types of interchanges. A common functionality is to allow an elevated crossover between an arterial road (upper) and free-way (below), where vehicles may switch between both of them using *slip roads*. The *diamond interchange* exactly achieves this. The main purpose is to allow a high speed continuous traffic flow without traffic signals. In large cities complex interchanges may even allow an elevated crossover of between two free-ways. The multi-level stack interchange, Figure 1.1, achieves this by means of several roads splitting and merging together.

FIGURE 1.1: A multi-level stack interchange located in Shanghai, China.

2. An *intersection* is an *at-grade* junction. In most situations two roads meet at a common point forming either a three-way (T junction) or four-way intersection. Naturally the traffic flow of such roads is at conflict and requires traffic signals to avoid collisions. Although interchanges eliminate several needs for traffic signals they are comparatively very expensive in terms of instalment and maintenance with respect to intersections.

   A *roundabout* is an alternative means of at-grade junction. The structure imposes a one-way direction, such that one must always give way to traffic approaching from that direction before entering the roundabout. Generally, no traffic signals are required but exceptions may exist. A general downside is the spatial requirements and low speeds (although the latter decreases the risks of severe accidents from occurring).

In summary interchanges and roundabouts are quite similar in various aspects but it is very difficult to alter their functionalities compared to traffic signals that require mostly software regulations only.

A *traffic signal* is a signalized device governing the traffic flow through an intersection by means of assigning the green light to different lanes. An *upstream* (*downstream*) lane leads traffic towards (away from) an intersection, thus a traffic signal must control upstream lanes in terms of vehicles, bicycles and pedestrians[1]. The road segment leading towards an intersection is commonly widened to include additional vehicle upstream lanes especially in terms of exclusive left and/or right turns. In addition a reverse procedure is usually adopted for the downstream lanes, i.e., at least two lanes merge together thus narrowing the road correspondingly.

[1]Pedestrian lanes are usually two-way directed with both directions obeying the same signal.

Traffic signals utilize different kinds of agreed color codes in order to control each type of upstream lane. The following is a widely used three-color code specifically for vehicle upstream lanes that operates with respect to a well-known four-step cyclic *display sequence*.

1. The red light displayed on its own forbids vehicles to proceed.
2. The red and yellow light displayed simultaneously (pre-green) hint that the signal is about to change to the green light. The step prepares waiting vehicles, in particular the leading one, to accelerate as to minimize any wasted green light.
3. The green light (*right of way*) displayed on its own allows the vehicles to proceed if safe to do so (an elaborated explanation of safe passage is given later on).
4. The yellow light displayed on its own alerts that the signal is about to change to the red light. The general rule suggests drivers to stop before the intersection stop line if safe to do so (in the presence of sufficient brake length and no close-up trailing vehicles), else they may proceed through. The *decision zone* refers to the time interval in which a driver is much unsure whether to proceed or stop which may cause a dangerous situations.

Each display has a duration denoted R, O, G, Y, respectively. A single iteration of every display is referred to as a *cycle*.Due to the cyclic property of a display sequence the starting point may be chosen arbitrarily. Note, the durations R, O, G, Y may be expressed either as a ratio or time interval of a cycle. This type of display sequence usually also handles bicycle upstream lanes. For pedestrians, signals usually switch between displaying either the red or green light on its own in which case the durations of step 1. and 3. are appropriately extended.

Before digging into the terminology of traffic signal optimization it should be noted that a traffic signal is an essential part of a traffic control system.

1.2.1 Traffic control system

A *traffic control system* (TCS) refers to any management tool attempting to improve traffic with respect to on one or several optimization criteria of a traffic network. It is comprised of numerous components and considerations that will be described below with respect to traffic signals.

- An *operational environment* is any relevant physical attribute of an intersection. These include the number of intersecting roads (e.g. T or fourway junction), the number of upstream and downstream lanes, the number of (and locations of) available traffic signals and its geometric properties (an uphill intersection may lead to reduced eyesight between opposing movements).
- A *traffic user* is any traffic entity such as passenger cars, transit buses, bicycles and pedestrians that is affected by the outcome of traffic signals. Therefore, it is important to consider various *traffic priorities* under prevailing conditions e.g. some types of entities should receive extended or shortened green light durations than others.
- An *operational objective* is an optimization criteria to be served by the TCS. They may vary depending on the desired outcomes of those that are responsible for

the TCS e.g. various municipalities. Some major objectives with respect to traffic signals include vehicle mobility and environmental impact.

- A *performance measure* is a quantification of the degree to which the operational objectives are fulfilled, thus it is what concludes whether or not a TCS is regarded optimal, or produces an approximate solution only. Corresponding examples for vehicle mobility and environmental impacts might be the maximization of vehicle throughput and minimization of vehicle stops, respectively.
- A *control measure* is the physical/digital infrastructure with regards to traffic signals that carry out the signal settings determined by a control strategy described below. Traffic signals are some of the most common control measures due their flexibility in terms of display output.
- A *control strategy* determines the signal settings based on the following characteristics of which the first two describe if a signal operates independently or within a signal network.
  - An *isolated* control strategy decides on the instructions of a single signal. This leads to a local optimization of the traffic network, however, possibly at the expense of any neighbouring traffic, as the instructions may be out of sync to other nearby signals.
  - A *coordinated* control strategy decides on the settings mainly of several consecutive signals along an arterial road. The coordination allows to improve on a greater subset of the traffic network.

  In addition, the characterization may be based on whether or not the proposed instructions are reconsidered.

  - A *fixed-time* control strategy determines the instructions based on traffic data obtained by detectors (e.g. induction loops and cameras) during some time horizon. Given historical data one may examine the distribution of traffic demands and propose corresponding settings, but it is assumed that the distribution in the future does not change (at least significantly). This is also referred to as a *static* control strategy. In its most extreme case once the optimal settings with respect to some distribution have been found they are never reconsidered. In reality, data is almost cost-free to collect once the detectors have been set up. Thus, the signal settings should be updated on a periodically basis. It is common to adopt a set of standard settings for specific times of a day, e.g. morning/afternoon rush hour and a third setting for any time else. This type of control strategy imposes little to no time constraints to the running time (efficiency) of any devoted optimization algorithm as the instructions are decided a long time in advance prior to be deployed. Therefore, additional attention may be drawn in developing effective algorithms, i.e., they produce the exact intended/expected results.
  - A *traffic-responsive* control strategy regularly examines real-time measurements of traffic demands as to adjust the instructions in an adaptive manner destined to serve the current traffic conditions more accurately. This is also known as a *dynamic* control strategy. This type of strategy, compared to the static approach, inherently imposes tighter time constraints on the algorithm running time, as the adjustments usually are required within

a short amount of time. Consequently, efficiency is often prioritized in terms of heuristics producing approximately optimal instructions only, however, in a reasonable amount of time. The degree of tighter time constraints depends on how the real-time measurements are to be utilized. In this sense one may distinguish between applying one of the following two approaches.

  - A *reactive* approach focuses on eliminating problems after they have happened. In the context of a control strategy the instructions should be adjusted almost instantaneously in response to current changing demands. This, however, might only resolve the issues for a very short amount of time, thus to maintain accurate handling of traffic demands a reactive approach must be applied frequently at a high rate with each reactive decision to be determined within an equally short amount of time.
  - A *proactive* approach utilize predictions to eliminate problems prior to their occurrence. The control strategy assumes that future demands for an extended period of time can be accurately predicted based on historical and current demands, and the instructions need to be determined for that entire period also. Although this is a computationally intensive task, it may successfully prevent certain traffic issues from happening, that would have appeared otherwise in a reactive approach.

These four control strategies work in pairs resulting in four different types of control strategies. In general, a coordinated traffic-responsive strategy is superior in terms of flexibility and potential improvements at the expense of high complexity, whereas the opposite holds for a isolated fixed-time strategy. Nonetheless, an isolated fixed-time strategy often proves sufficient thus encouraging the consideration of traffic-responsive strategies only, if the possibilities of less complicated approaches have been fully explored.

- An *evaluation strategy* carries out the assessment of the TCS by determining the performance measures that determine the degree of optimality of the TCS with respect to the operational objectives. Ideally, the performance measures are derived analytically given a complete knowledge of traffic conditions and traffic signal instructions at any point in time. In reality, this may be feasible for simple cases, i.e., a single traffic signal in conjunction with assumed simplified constant traffic conditions, however, the complexity significantly increases in dealing with coordinated signals such that the performance measures must be estimated by means of traffic simulations. There are two main approaches to traffic simulation.

  1. A *microscopic* simulation considers each traffic entity individually by examining microscopic properties like its position and velocity. This allows accurate estimation of a variety of quantitative measurements such as the number of vehicle stops or time spent in the network.
  2. A *macroscopic* simulation considers traffic as a single entity in terms of a fluid stream, which is characterized by macroscopic properties such as average speed or density. This makes it easier to estimate measurements such as the throughput of vehicles by a single number, however, less accurate than the microscopic model due to the reliance on averages only.

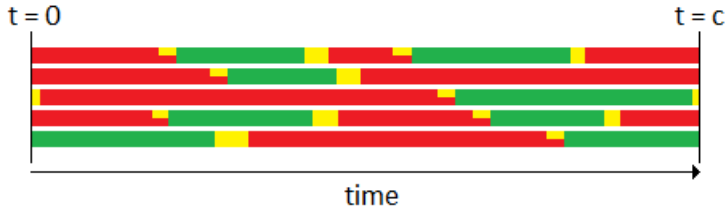


FIGURE 1.2: Five phases sharing some joint cycle length.

In both methodologies the key point is to derive/estimate similar performance measures with respect to a non-optimized equivalent scenario for comparison, in order to quantify the degree of optimality. If the improvements turn out significant then one should consider implementing the solution in real life.

1.2.2 Traffic signal optimization

We present the definitions necessary for the optimization of traffic signals.

- A *movement* is the smallest set of upstream lanes that is distinguishable by a set of similar physical signals, such that each upstream lane obey the exact same display sequence of those signals. Therefore the number of movements does not necessarily equal the number of upstream lanes. A *divided* (*undivided*) movement is a movement whose traffic may proceed in a single direction only (several directions).

- A *phase* is any non-empty subset of movements such that the signals of each movement cycle through a joint display sequence.

- The *split* is the green duration, G, of a cycle but one may also include the yellow duration, Y.

- The *cycle length*, c, is the required time span of a single cycle iteration such that any phase has gone through its display sequence exactly once. Note that $c = R + O + G + Y$ for each phase. It should be noted, however, that one may consider a *double cycle* such that a phase may repeat its display sequence twice during some global cycle length whilst other phases perform a single iteration of their display sequences only. In general, any phase may go through any number of arbitrary display sequences during some time horizon, however all phases at some point must sync with respect to a joint cycle length to ensure periodicity, see Figure 1.2.

- An *offset* is the relative timings among the phases of coordinated intersections. It is defined only when dealing with a coordinated control strategy.

A *stage* is often mentioned in the literature but without care, thus we avoid that term entirely **reformulate this statement**

We describe several tasks that need be determined based on the above definitions for which there are several considerations that must be taken into account.

- A *phase specification* is the set of decided compositions of phases, given a set of movements, that must be served at an intersection. An apparent issue is that the movements composing a phase may conflict with each other if the trajectories inside the intersection followed by the users of those movements intersect. Figure 1.3 [7] illustrates all possible conflict points in a four-way intersection

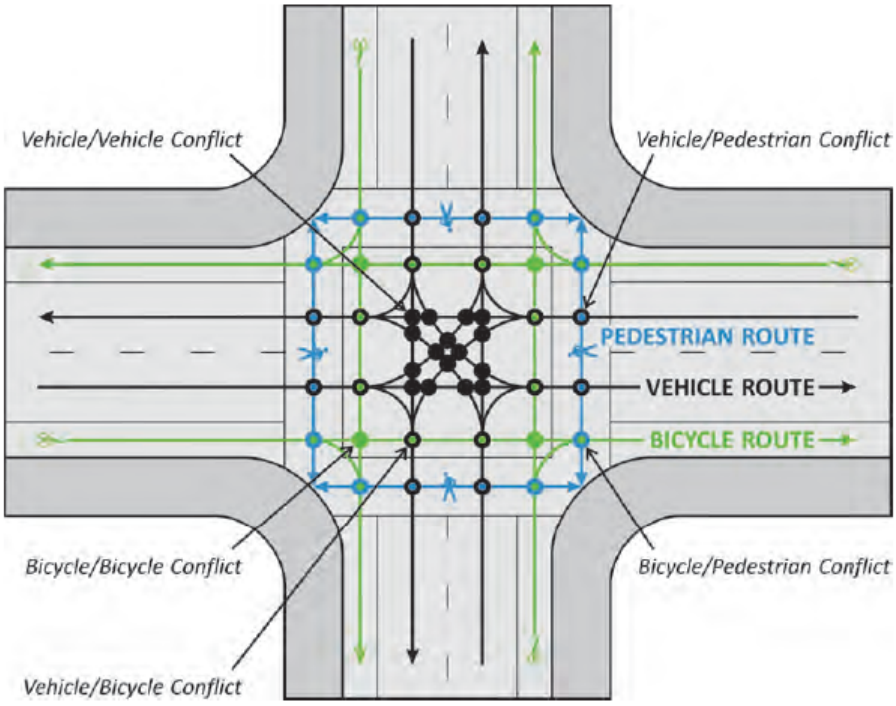


FIGURE 1.3: Illustration of every possible conflict point in a four-way intersection with the specified physical attributes.

among all traffic users. Note that some conflict points could be avoided in the presence of multiple downstream lanes, for instance the south-north and east-north vehicle movements would not conflict if the northbound downstream road contained two separate lanes to collect traffic for each movements.

Two movements within a phase are mutually *protected* if their intersection trajectories do not intersect at a conflict point, otherwise one of them has to give way to the other such that the former is referred to as a *permitted* movement. A commonly agreed order of priorities suggests that vehicles must give way to bicycles and pedestrians, and bicycles must give way to pedestrians. In addition, given two conflicting movements between equivalent traffic users the ones making a turn should always give way to those proceeding straight forward. Clearly, there are instances in which both movements have equal priority and therefore should not appear in a common phase. Protected movements generally prioritize safety but typically require more phases in order to serve all movements which increases the lost time required when transitioning between phases. The opposite implications apply to permitted movements.

An useful alternative description of the above cases is as follows. A pair of movements within a phase may be *compatible* - they are mutually protected), *semi-compatible* - there exists an obvious order of priority such that one movement is protected and the other permitted) or *incompatible* - there exists no such order).

It should be emphasized that the exact jurisdictions of an order of priorities may vary depending on the circumstances of the intersection, i.e., rules dictated by a municipality (or whoever being responsible for the traffic signal). For instance a traffic engineer may adopt a phase specification that signifies

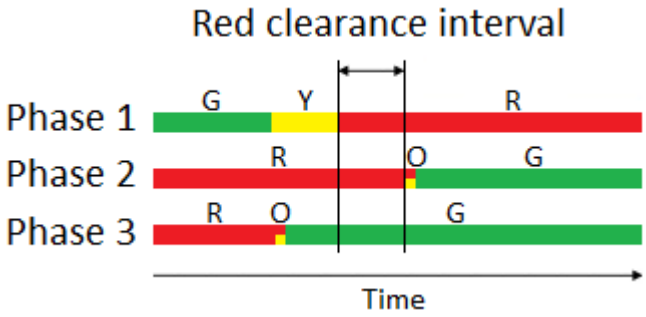


FIGURE 1.4: The red clearance interval is required for the safe transition between two incompatible phases P1 and P2. A third phase P3 is shown to clarify, that the red clearance interval is not necessarily a red period for all phases since P3 is compatible with both P1 and P2.

certain risks in terms of safety in the favour of other objectives, while other may strive to always maximize safety.

The purpose of a phase specification is to simplify the the signal settings by grouping together certain movements that, intuitively speaking, should be timed simultaneously. However, the simplicity may disregard (at the expense of increased complexity) better solutions.

- A *phase schedule* is a set of start and stop times for the split of each phase within a phase specification with respect to some reference time-stamp that defines the relative positions of the splits among all phases.In scheduling the splits one must consider the compatibilities among phases with respect to overlapping splits. These are as follows (an extension of the definition for movements)
  - Two phases are *compatible* if the union of their movements are mutually protected, thus their splits may overlap at any given time.
  - Two phases are *semi-compatible* if there exists an agreed order of priority between each pair of movements from the union of the phases such that one movement is protected and the other permitted. In this case the splits of the phases may overlap, however, potentially subjected to additional constraints.
  - Two phases are incompatible if there exists at least one pair of movements from the union of the phases for which there exists no agreed order of priority, thus the split of these phases must not overlap. Incompatible phases impose most restrictions on the phase scheduling. In order to safely transit between two incompatible phases, $\Phi_1 \rightarrow \Phi_2$ one considers the following two intervals.
    * The *yellow change interval* **is** the yellow display duration, Y, whose function is to warn that the current phase is about to end.
    * The *red clearance interval* is a, mostly optional, joint short period associated with R for Φ_1 (in the beginning of R) and Φ_2 (at the end of R), see Figure 1.4. The purpose is to provide enough time to clear any traffic in the intersection resulting from Φ_1 that would interfere at the conflict point with Φ_2. This is to resolve the issue of drivers unable to safely stop during Y.

- The split of a phase must be maximized in order to accommodate the demand for each movement composing the phase. The *minimum green time* (*maximum green time*) is the minimum (maximum) split that can be assigned to a phase, but their applications slightly varies depending on the control strategy in question.
  - For fixed-time only the minimum duration is relevant that must reflect e.g. the slowest moving pedestrians and general vehicle expectations.
  - For traffic-responsive the minimum duration prevents a currently active phase from being terminated despite an excessive queue in conflicting phases. In contrast, the maximum duration exactly prevents a currently active phase with excessive queue to maintain the right of way despite only a few vehicles waiting in conflicting phases.

  The splits also requires a predetermined phase specification or they may be decided concurrently.

- Given undersaturated traffic flows the cycle length must be minimized. However, given a cycle length being too short the significance of the ratio between hourly green and hourly lost time becomes infeasible.

  In the context of traffic-responsive control strategies the cycle length is of minor importance since phases may be prioritized randomly in response to randomly changing demands.

- The offset must be determined to allow green waves such that vehicles may pass consecutive intersections without the need to stop. Coordination may be either one - or two-way along arterial roads. One-way coordination may be easily configured given fixed stages for each intersection as one must only take into account the expected travel time between intersections. Two-way coordination inherently is more difficult
  - The intersections in question must have a joint cycle length with respect to long term coordination.
  - Consider two intersections for which a two-way coordination is desired. The offset is given by

    $$\frac{L}{v} = n\frac{c}{2} \tag{1.1}$$

    where
    - L is the distance between the intersections,
    - v is the average velocity of the vehicle platoons in both directions,
    - and $n \in \mathbb{N}$ is the integer multiple between cycles.

  Generally, the main issue is the integer multiple of half a cycle length due to the fact that the speed limits in each direction must be equal[2].

The *ring-and-barrier concept* is a helpful means of illustrating the organization of phase orderings with respect to splits. Several examples with respect to a four-way intersection are found in [7]. Figure 1.5 [7] illustrates the most simple example. A *ring* is a sequence of incompatible phases with respect to some order of priorities. A

[2]We believe this certainly is true everywhere in the world.

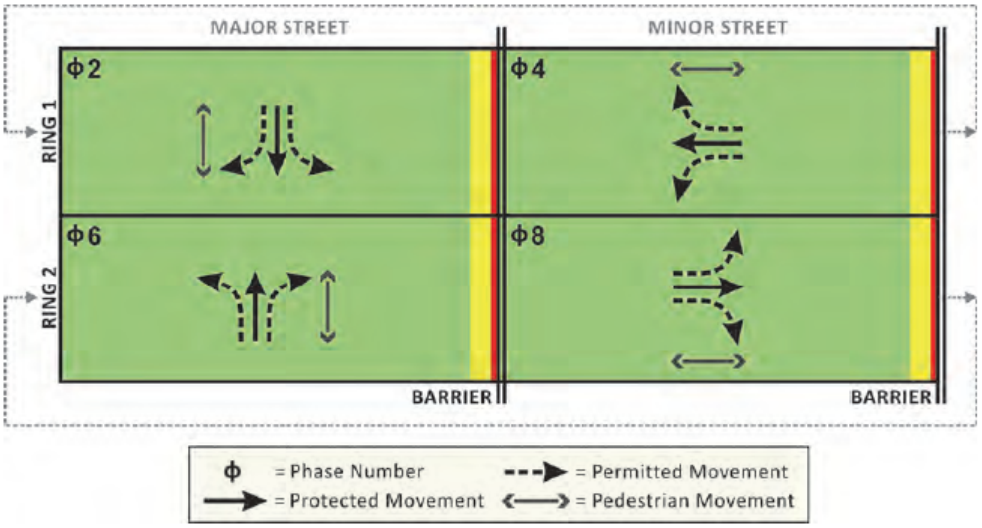


FIGURE 1.5: **ADD CAPTION**

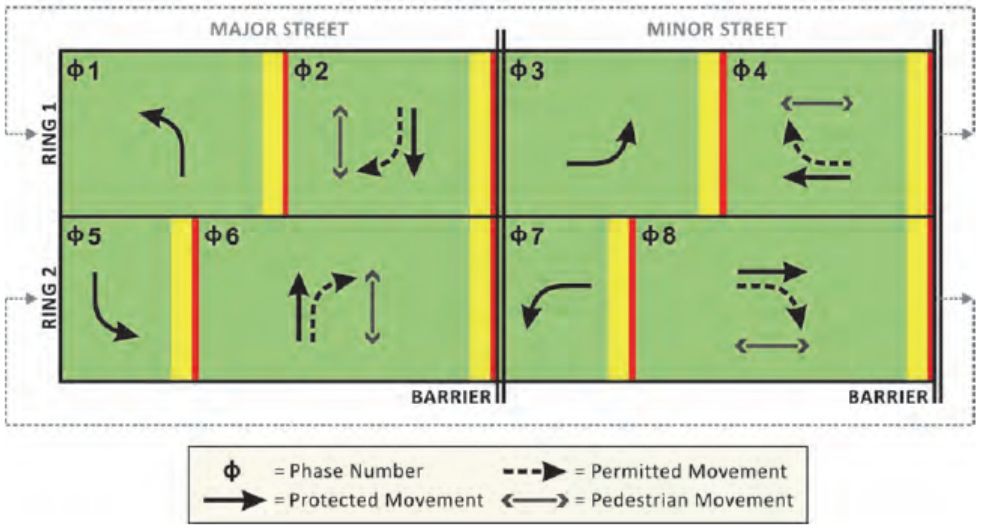


FIGURE 1.6: **ADD CAPTION**

barrier divides the rings in such a way that any phase in ring 1 between two barriers may time with any phase in ring 2 (compatible or semi-compatible) between the same two barriers and vice versa. These possibilities are more evident from the example in Figure 1.6 [7] which also illustrates how one may vary the splits of the phases. For instance, Φ1 may give more time to Φ2, and/or Φ8 should be skipped entirely in favour of phase Φ7 (assuming a minimum green time equal to 0). In addition, a barrier also marks the point in time at which the phases in both rings must end simultaneously. In the examples, these markings divide the rings with respect to the major and minor street, commonly the arterial road and small streets, respectively. Lastly, the entire ring-and-barrier amounts to a single cycle.

A *time-space diagram* is another helpful means of visualization, in the context of coordinated traffic signals. Basically, one considers an arterial road along consecutive intersections, see Figure 1.7. The objective is to minimize the number of vehicle stops by the coordination of "green waves". The *vehicle trajectory bandwidth* is the horizontal distance between the first and last vehicle trajectory that should be maximized in order to grant the green wave to as many vehicles as possible. This is the

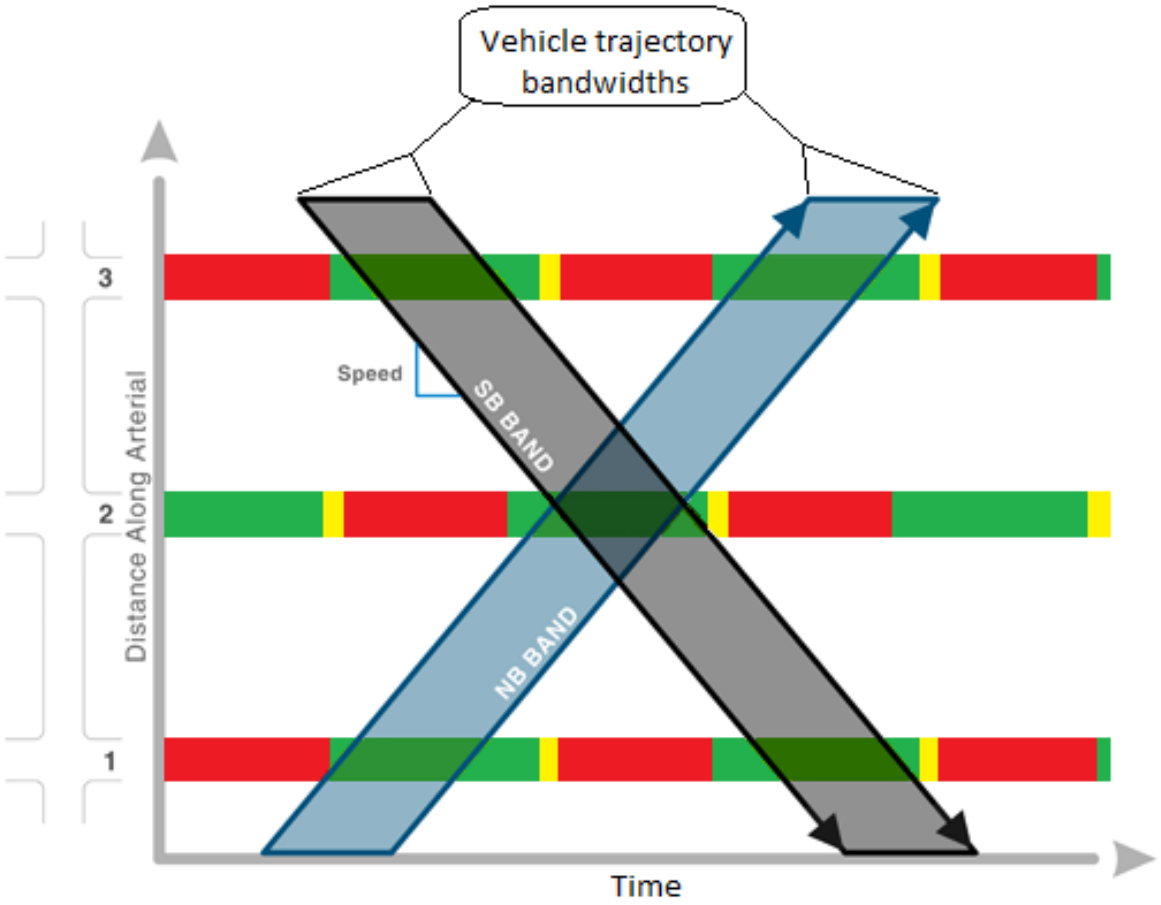


FIGURE 1.7: Time-space diagram for coordination of three consecutive intersections. The south and north bound bands consists of several vehicle trajectories.

key idea of the method *MAXBAND* to be explained in the next chapter.

1.2.3 Other terms

Headway is a central term when describing traffic flow. There are various definitions depending on the situation in question. Mostly, it refers to the distance (tip to tip, or tip to tail) between consecutive vehicles or the time taken for the trailing vehicle to cover that distance. Large headways, especially on freeways, increase vehicle safety - a rule of thump is to maintain a distance behind the leading car being greater than the distance covered during the average reaction time in case of an emergency braking of the leading vehicle, but at the expense of poorly utilized road capacity. The exact opposite holds for low headways. In *transit* traffic the term has different meanings and heavily relates to the delay of transit vehicles. A set of buses serving a common route, each one punctual on time-schedule, implies an optimal headway, whereas a bus behind its schedule may render the trailing one redundant. This is a serious flaw in resource allocations: a late bus may serve a stop station, leaving the trailing one empty.

A *platoon* is a line of vehicles having in between very small headways, that are accomplished by means of electronic (possibly mechanical) coupling. The line of vehicles then acts as a single unit controlled by the leading vehicle. Some advantages include simultaneous acceleration and braking, thus nullifying the downsides of

human reaction time, and reduced air resistance (thus less fuel consumption) due to *drafting*. The concept may be regarded as intending to increase the capacity of roads by means of bus imitations. As of today it remains mostly a proposed improvement of traffic flow only, as actual realizations experience several challenges, in particular maximizing the reliability of autonomous vehicle systems. Of course, platoons may still occur spontaneously, although with significantly larger headways due to human reaction time. They are essential when determining if the coordination of traffic signal intersections would be beneficial [7], i.e., if large platoons approaching a series of consecutive intersections should be prioritized as to minimize the number of vehicle stops.

A *passenger car unit* (pcu) is a metric used in quantifying the load or stress that any traffic entity has on traffic in general compared to a single passenger car. An *equivalence factor* is assigned to each entity. For instance, a truck may have a factor of 3 due to poor acceleration and spatial requirements, meaning that it is equivalent to three pcus. In addition, the factor of a bicycle may be as small as 0.2 meaning it imposes $1/5$ the load or stress of a single pcu. Note, a bus, although having similar drawbacks as the above truck, may contain a large amount of people relative to its size meaning that the traffic is relieved by an equally large amount of pcus, thus considerably decreasing its equivalence factor. Alternatively, the bus is empty, thus posing a huge load on the traffic. The point is that the equivalence factor for a vehicle may vary with time.

A more accurate metric, especially in the context of traffic signal intersections, is the *through car unit* (tcu) that builds upon the assumptions for a pcu. The metric takes into account that pcus proceeding straight forward imposes more throughput compared to turning pcus as the latter require lower speeds to perform a turn.

The following technical terms are associated with an intersection movement.

- The *saturation headway* is the minimum time elapsed between the tip of two consecutive pcus passing a common point. Clearly, this implies that the pcus maintain the corresponding physical minimum safe tip-tail headway.
- The *saturation flow rate* (pcus per hour of green) of an intersection movement is the maximum number of pcus in an infinite platoon that may enter the intersection given one hour of green. It is computed by $\frac{3600}{h}$ where h is the saturation headway.
- In *free flow* the flow rate is at most the saturation flow rate.
- A *standing queue* is the queue of stationary pcus that is present as signal turns green. It does not account for pcus arriving at the back of the queue after the signal has turned green.
- To define the remaining terms we first need to consider the various characteristic periods within a display sequence, see Figure 1.8.
  - The *effective green* g (seconds) is the portion of $G + Y$ where pcus enter the intersection in a stable moving platoon at the saturation flow rate. One may also express g in terms of a portion of a cycle, g/c, known as the *effective green ratio*.
  - The *start-up delay* ℓ_1 is equal to the sum of headway delays Δi compared to the saturation headway, see Figure 1.9 [3]. A standing queue receiving the green light is discharging pcus at the saturation flow rate when a vehicle crosses the stop line at the speed limit because the following vehicles will

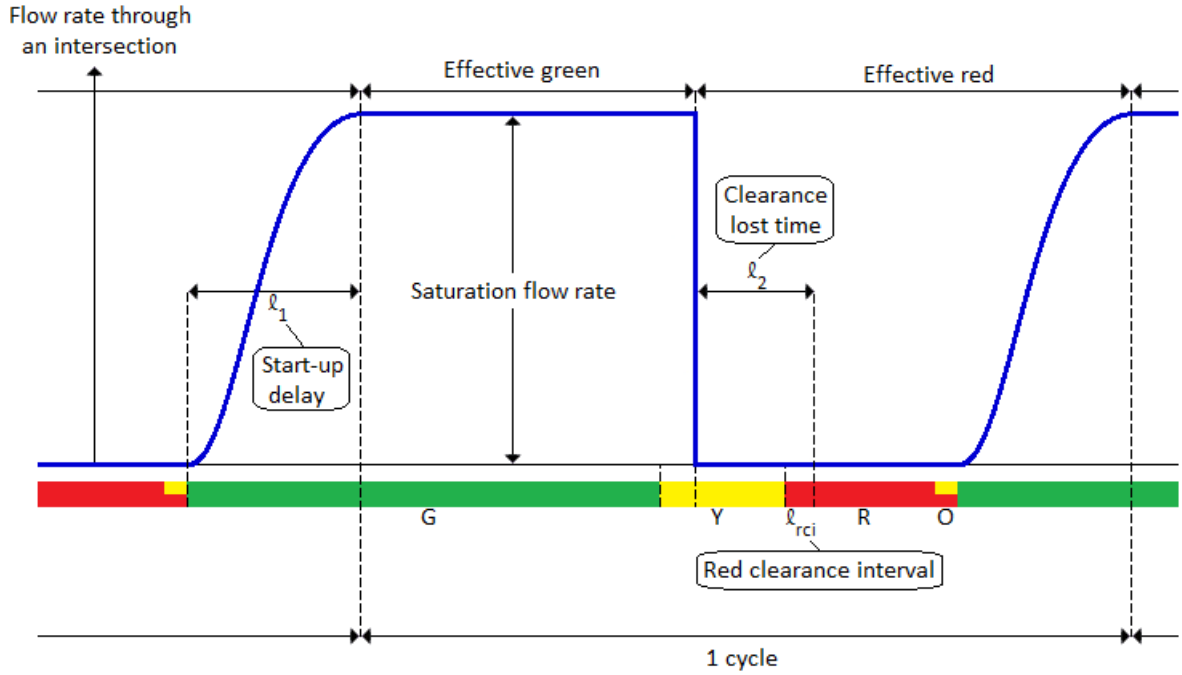


FIGURE 1.8: Intersection flow rate of some phase as a function of its durations in a display sequence.

also be at the speed limit and they will maintain the saturation headway. As a rule of thump this is achieved by the 4th or 5th vehicle.

The *clearance lost time* ℓ_2 is equal to the red clearance interval plus the portion of Y in which no pcus enters the intersection as they are able to stop safely (the opposite holds for the remainder of Y).

The *lost time* ℓ (seconds) is the time duration in which the intersection is not effectively utilized to serve pcus at the saturation flow rate due to a change between conflicting phases. It is equal to the sum of start-up delay and clearance lost time. In Figure 1.8 it is important to note that the start-up delay is not equal to the portion of G as specified, rather it indicates the time interval of G in which the pcus are not entering the intersection at the saturation flow rate.

– The *effective red* r (seconds) is the portion of the cycle in which pcus do not enter the intersection in a stable moving platoon at the saturation flow rate, however, in the majority of the interval pcus are stationary.

When deciding how much green time to allocate for a phase it is the effective green, not G, being of interest. Formally, given g, the above concepts are related to one another as follows.

$$\ell = \ell_1 + \ell_2 = G + Y - g + \ell_{rci} \tag{1.2}$$

$$c = r + g \tag{1.3}$$

$$c = R + O + G + Y \tag{1.4}$$

- The *capacity* (pcus per cycle) is defined by the product of the effective green ratio and the saturation flow rate.

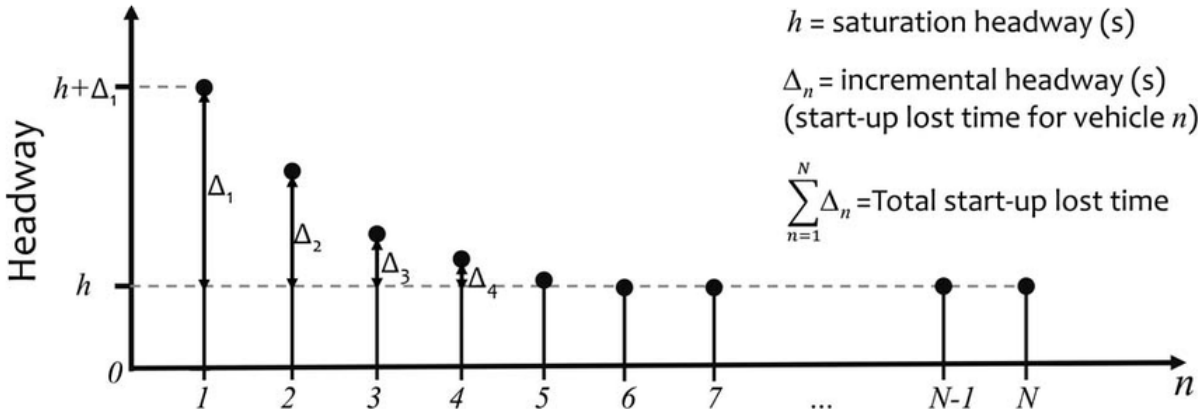


FIGURE 1.9: The x-axis shows the vehicle index in a standing queue containing N vehicles. The y-axis shows the time headway

- The *arrival rate* (pcus per hour) is the number of pcus intending to cross the intersection using the movement during one hour.
- The *demand* (pcus per cycle) is given by the product of the arrival rate and the cycle time.
- The *degree of saturation* is the ratio between demand and capacity.
- Traffic *volume* (pcus per cycle) is the actual number of pcus that are able to cross the intersection in a cycle.
- A movement suffers from (*oversaturation*) if being unable to fully discharge all pcus in its standing queue during a cycle. Demand and volume are proportional until the capacity of the movement is exceeded in which case the volume remains constant but the demand may still increase, thus leading to long term queue build ups.
- A *phase failure* is the occurrence of at least one pcu in a standing queue being unable to cross the intersection during a cycle, thus leading to long term queue build ups.

1.3 Literature review

As already mentioned control strategies are characterized roughly by four operating scenarios. Important methodologies for each of them with respect to traffic signals are highlighted in [8]. We will describe each methodology with respect to the terminology given in this chapter.

- Isolated fixed-time: The methods *SIGSET* and *SIGCAP* assume predetermined phase specifications in order to find the optimal splits and cycle length. The phase scheduling does not affect the main objectives as the phases are served sequentially. Nonetheless, the methods may be stated to determine the stage specification problem. Given n phases the methods specify the splits $\lambda_1, ..., \lambda_n \in [0, 1]$ as ratios of the cycle length c such that

$$\lambda_0 + \lambda_1 + ... + \lambda_n = 1 \tag{1.5}$$

where $\lambda_0 = \frac{\sum_{i=1}^{n} \ell_i^{rci}}{c}$ from which the insignificance of the phase orderings becomes apparent.

The method presented in [4] solves a similar problem, with the addition of determining the optimal phase orderings as the phases may not be mutually incompatible. The model rises a binary-mixed-integer-linear (MILP) programming problem, which is solved through a branch and bound method utilizing the Simplex algorithm. In the following chapter we explore the method in details with several proposed improvements.

- Coordinated fixed-time: *MAXBAND* attempts to maximize the number of green waves to reduce vehicles stops. In other words it determines appropriate offsets. The concept is similar to the *time-space diagram* described in [7]. The method requires a binary-mixed-integer-linear-programming problem to be solved.

  TRANSYT employs a local search heuristic. Given input decision variables (splits, offsets and cycle time) the algorithm introduces small local neighbourhood changes in order to improve the operational objective, until a local minimum has been found.

- Isolated traffic-responsive: The *vehicle interval method* is a two-phase based method. Its basic principle is to determine, for each interval, whether the current phase should end in favour of the other. This is achieved by means of loop inductors located some distance away from the intersection.

  Sameh Samra et al. [9] considers a problem referred to as the *traffic control problem*. It assumes a set $\mathcal{K}$ of phase specifications with the phases mutually incompatible as with *SIGSET* and *SIGCAP*. Formally, it asks to determine a timing plan $\Omega = \langle (K_0, t_0), ..., (K_i, t_i) \rangle$ where $K_i \in \mathcal{K}$ and $t_i < T$, i.e., a sequence of phases, each assigned a split within a time horizon T. Note that T is *not* a cycle length to be enforced since traffic-responsive strategies do not strictly cycle through a specified stage. The authors present a new linear time and space algorithm to solve the problem. It improves upon an existing one having complexity $O(T^3)$ in time and $O(T^2)$ in space. It utilizes a dynamic programming formulation of the problem that heavily relies on predictions of demands, therefore being proactive. Simulation studies were conducted in confirming the superiority of the algorithm compared to other modern methodologies including reinforcement learning, genetic programming and neural networks. The accuracy of the predictions of the demands given to the algorithm, however, were assumed to be perfect.

- Coordinated traffic-responsive: The method *SCOOT* basically applies the same methodology as TRANSYT, with respect to real-time measurement, however.

  Model-Based Optimization Methods like *OPAC*, *PRODYN*, *CRONOS*, and *RHODES* solve in real time a dynamic optimization problem based on a sampling time of 2-5 seconds of real-time traffic measurements. Generally the problems presented require exponential computation time, thus leaving the problem infeasible for more than a single intersection.

The title of this thesis itself is a broad area of research largely depending on the operational objectives in question. For instance the convenience of and demands for private cars and the feeling of independence may be challenged by introducing transit vehicle priorities by increasing the likelyhood of time-schedule punctualities.

As a matter of fact a single bus has the potential in replacing 15-20 cars. A comprehensive review of transit priority at traffic signal intersections is given in [5].

Chapter 2

Modeling

In the previous chapter we defined the following optimization tasks: 1) phase specification, 2) phase scheduling, 3) splits, 4) cycle length and 5) offsets. In the following we describe current existing methods and our proposed extensions thereof with respect to each characteristic of a control strategy.

2.1 Fixed-time isolated signal

2.1.1 SIGSET and SIGCAP

SIGSET [2] and SIGCAP [1] both assume a predetermined phase specification $\mathcal{K} = \{K_1, ..., K_n\}$ based on the set of movements, $H = \{h_1, ..., h_m\}$, of the intersection in question such that $K_i \subseteq H$ for $i = 1..n$. Clearly, it must hold that

$$\bigcup_{K_i \in \mathcal{K}} K_i = H.$$

however, both methods do not require that

$$\bigcap_{K_i \in \mathcal{K}} K_i = \emptyset$$

thus, phases may share a common movement such that it receives a green duration contribution from multiple phases. The phase splits are treated as ratios relative to the cycle length, c, in order to decide on both tasks. In considering the split ratios $\lambda_1, ..., \lambda_n \in [0, 1]$ it must hold that

$$\lambda_0 + \lambda_1 + ... + \lambda_n = 1 \tag{2.1}$$

where $\lambda_0 = \frac{\sum_{i=1}^{n} \ell_i^{rci}}{c}$ is the total lost time during a single cycle. A lower and upper bound on c must be imposed. Basically, each λ_i are assigned with respect to the green time movement requirements. This is achieved by the following main constraints to avoid queue building

$$s_j \sum_{i=1}^{n} a_{ij} \lambda_i \geq q_j \qquad \forall h_j \in H \tag{2.2}$$

where q_j is the arrival rate and s_j the saturation flow rate of some movement $h_j \in H$, and $a_{ij} \in \{0, 1\}$ depending whether or not $h_j \in K_i$ for some $K_i \in \mathcal{K}$. Finally, it should be noted that the phase schedule is redundant as the phases are served sequentially, i.e., mutually incompatible, meaning that every red clearance interval is equivalent to an all-red duration. Overall the disadvantage of these methods is

due to a fixed predetermined phase specification that also dictates the inability to consider different phase schedules.

Both methods are partly dominated by G. Improta and G.E. Cantarella [4], (IC), on which the modelling principles of subsequent sections are based with respect to a fixed-time isolated signal.

2.1.2 The model (IC)

The model (IC) decides the phase specification, phase schedule, splits and cycle length, thus signifying an apparent advantage compared to SIGSET and SIGCAP. In contrast, however, it requires that

$$\bigcap_{K_i \in \mathcal{K}} K_i = \emptyset$$

that may lead to other (possibly inferior) optimal solutions. The phase specification is decided in considering only the movements (and the compatibility relations among them) of the intersection in question. Thus, the outcome depends on the priority order among the movements stated by those responsible for the traffic signal as pointed out in Chapter 1. The solution can be visualized by drawing the display sequence for each movement, from which the alignments of the splits reveal the phase specifications (a single movement also accounts for a single phase). The movements may be served along side each other (non-sequentially), thus the impact of the phase schedule (more accurately in this case the movement schedule) must be considered. Finally, the splits and cycle length are both decided by means of ratios (similar to SIGSET and SIGCAP) and absolute durations.

2.1.3 An extended model of (IC)

The model to be presented includes various extensions of (IC) that is briefly summarized in the following and explained in details in subsequent sections.

- In contrast to the model (IC), this model in scheduling the splits of the movements considers semi-compatibilities among movements and potential reduced movement saturation flow rates. Such movements may increase the intersection flow rate given undersaturated conditions, but at the expense of safety regarding drivers determination of available safe gaps. The objective of semi-compatible movements is to investigate whether they offer any potential benefits with respect to various degrees of saturation.

- **This idea has yet to be formalized in the model** A potential drawback of (IC) is the fact that each movement has only a single split during a single cycle iteration (in contrast to SIGSET and SIGCAP). This can be overcome by introducing multiple identical movements with individual splits to be scheduled. In order for the idea to be incorporated we need to consider the following.

  – The splits of two identical movements should not overlap. To see this assume on the contrary two overlapping splits. If a split completely covers the other then the latter is redundant. In addition, two partially overlapping splits make up a single split, however that would imply that either split by itself could account for that single split thus also leaving the other redundant.

  - Each additionally introduced movement should require a minimum effective green ratio similar to that of the original movement.
  - The number of additionally introduced movements depends on the minimum effective green ratio, $\underline{g}$. Given an upper bound on the cycle length (if c is varying), $\bar{c}$, it is possible to schedule at most $\lfloor \bar{c}/\underline{g} \rfloor - 1$ additional movements. However, the total number of movements may render the problem computationally infeasible.

As in (IC) we consider the elimination of the cycle length variable, c, by re-expressing relevant quantities in terms of ratios of c. A new variable, $z = 1/c$, is introduced that is bounded by any desired lower and upper values for c. That way, the cycle length can be either predetermined as a parameter in specifying an equal lower and upper bound, or treated as a variable.

In the following model some constraints are presented with respect to desired implementation features of the open source optimization solver `GUROBI`.

2.1.3.1 Input parameters

- Let $H = \{h_1, ..., h_m\}$ be the set of movements of the intersection in question.
- Define a partial order (order of priority) on H such that $h_i, h_j \in H$
  - are *incomparable* if they are compatible,
  - have equivalent priority, $h_i \equiv h_j$, if they are incompatible,
  - have a priority order, $h_i \prec h_j$ (h_i must give way to h_j), if they are semi-compatible.
- Define the sets

$$\begin{aligned} I &= \{(i,j) \in \{1..m\} \times \{1..m\} \mid h_i \equiv h_j\} \\ \mathcal{S} &= \{S_1, ..., S_m\} \\ \text{with } S_i &= \{j \in \{1..m\} \setminus \{i\} \mid h_i \prec h_j\} \qquad i = 1..m \end{aligned}$$

- Let $\ell_{rci}^{ij} \geq 0$ (seconds) be the red clearance interval for $(i,j) \in I$ when transitioning from h_i to h_j. Note that we may have $\ell_{rci}^{ij} \neq \ell_{rci}^{ji}$.
- For each movement $h_i \in H$ we define (all in seconds)
  - $\underline{g}_i$ to be its minimum effective green.
  - ℓ_i to be its lost time, i.e., the sum of start-up delay and clearance lost time. Note, in contrast to the definition in Chapter 1 the ℓ_i in this model does *not* include the red clearance interval, ℓ_{rci}^{ij}. Instead, it is specified independently, see above.
- Let $q_i, s_i \geq 0$ (both passenger car units per hour) be the arrival rate and saturation flow rate, respectively, for some movement $h_i \in H$.
- Let $\zeta_i = \dfrac{q_i}{s_i}$ be the flow ratio that restricts the required effective green ratio, g_i (to be defined), of some movement $h_i \in H$. The flow ratio is justified as follows. During the effective red ratio, $1 - g_i$, the expected standing queue build up per cycle is assumed to be $q_i(1 - g_i)/(3600/c)$. During g_i the passenger car units

are discharged at rate $s_i - q_i$ due to additional vehicle arrivals while clearing the standing queue such that $g_i(s_i - q_i)/(3600/c)$ is the number of severable passenger car units per cycle in movement h_i. Thus, to prevent oversaturation the growing standing queue must be cleared within each cycle, i.e.

$$\frac{q(1-g)}{3600/c}\left(\frac{s}{3600/c} - \frac{q}{3600/c}\right)^{-1} g = 1 \Leftrightarrow \tag{2.3}$$

$$q(1-q) = (s-q)g \Leftrightarrow \tag{2.4}$$

$$\frac{q}{s} = g \tag{2.5}$$

A traffic signal is not operating effectively when it displays green for a movement with free flow compared to when discharging a standing queue. This is due to potential standing queue build ups in every other movement, hence the equality sign in (2.41). Nevertheless, one should always allow a larger green ratio that does not affect the queue handling of every other movement, hence one should consider (2.41) be

$$g_i \geq \frac{q_i}{s_i} \tag{2.6}$$

In the constraint section we will introduce a constraint with respect to (2.42).

2.1.3.2 Decision variables

- For each movement $h_i \in H$ we define
  - g_i to be its effective green ratio.
  - u_i to be its starting time of G and v_i its ending time of Y such that $v_i - u_i$ is the split ratio of h_i that accounts for both g_i and ℓ_i.
- Let $f \geq 0$ be the intersection capacity factor. This is a maximum common multiple factor with respect to all movements.
- Let z be the reciprocal cycle length $\frac{1}{c}$ such that $\frac{1}{\bar{c}} \leq z \leq \frac{1}{\underline{c}}$ where $\underline{c}$ is the lower bound and $\bar{c}$ the upper bound (both seconds) on the cycle length c.

Moreover, we define the following auxiliary variables

- $w_{ij} \in \{0,1\}$ to be a binary variable defined for each $(i,j) \in I$ such that $w_{ij} = 0$ if the split of h_i precedes that of h_j and $w_{ij} = 1$ otherwise, i.e., the split of h_j precedes that of h_i.
- $x_{ij} \in \{0,1\}$ to be a binary variable defined for each $j \in S_i$ for $i = 1..m$ such that $x_{ij} = 0$ if the splits of h_i and h_j overlap and $x_{ij} = 1$ otherwise. The variable reflects the fact that two semi-compatible movements may overlap *but it is not required*.
- $y_{ij} \in \{0,1\}$ to be a binary variable defined for each $j \in S_i$ for $i = 1..m$ such that $y_{ij} = 0$ if the split of h_i precedes that of h_j and $y_{ij} = 1$ otherwise, i.e., the split of h_j precedes that of h_i.

2.1.3.3 Constraints

The minimum effective green ratio ensures a minimum flow rate

$$z\underline{g}_i \leq g_i \qquad \forall h_i \in H \tag{2.7}$$

Since $v_i - u_i = G + Y$, assuming G, Y are ratios also, it follows from (1.2) that

$$v_i - u_i = g_i + z\ell_i \qquad \forall h_i \in H \tag{2.8}$$

note, however, the absence of the red clearance interval which is specified independently.

Each movement $h_i \in H$ must have sufficient effective green ratio with respect to its flow ratio

$$g_i \geq f\zeta_i \qquad \forall h_i \in H \tag{2.9}$$

Here, f must be maximized since a high value indicates a sufficient utilization of the intersection capacity. If $f = 1$ then in theory g_i suffices for all movements assuming that pcus arrive uniformly, however, as this is unlikely to happen in real life the case of $f \geq 1$ is always desired. In contrast, it may not be possible to allocate a sufficient g_i due to a high value of ζ_i in which case we must have $f < 1$ in order to satisfy the constraint for the movement in question and the overall feasibility of the model. Finally, in theory a result with $f < 1$ causes oversaturation due to standing queue build ups resulting from phase failures.

A complete traffic signal plan, when implemented in physical signals, may be initiated starting from an arbitrary point in time due to its cyclic property. For simplicity we assume

$$u_i \geq 0 \qquad \forall h_i \in H \tag{2.10}$$

such that the resulting plan of this model should be initiated at $t = 0$. In addition, if $u_i \geq 1$ then u_i can be shifted 1 units backwards (1 cycle) along the time axis, thus

$$u_i \leq 1 \qquad \forall h_i \in H \tag{2.11}$$

The bounds on v_i are discussed later on.

Movement incompatibility constraints

The following constraints ensure that the split of two incompatible movements do not overlap, see Figure 2.1. We consider some $(i, j) \in I$. If the split of h_i precedes that of h_j, i.e., $w_{ij} = 0$ then

$$z\ell_{rci}^{ij} + v_i \leq u_j \tag{2.12}$$

$$z\ell_{rci}^{ji} + v_j \leq u_i + 1 \tag{2.13}$$

and vice versa

$$z\ell_{rci}^{ij} + v_i \leq u_j + 1 \tag{2.14}$$

$$z\ell_{rci}^{ji} + v_j \leq u_i \tag{2.15}$$

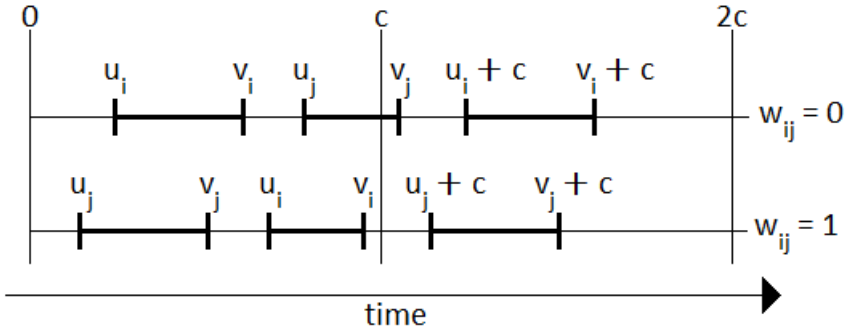


FIGURE 2.1: Time horizon for two cycles which amounts to a time span of two cycle lengths.

Constraints (2.12-2.14) and (2.13-2.15) can be combined, respectively, as follows

$$\left.\begin{aligned} z\ell_{rci}^{ij} + v_i &\leq w_{ij} + u_j \\ z\ell_{rci}^{ji} + v_j &\leq (1 - w_{ij}) + u_i \end{aligned}\right\} \forall (i,j) \in I \tag{2.16}$$

Movement semi-compatibility constraints

We consider some $j \in S_i$ for $S_i \in \mathcal{S}$. The splits $[u_i, v_i]$ and $[u_j, v_j]$ are permitted to overlap *but it is not required*. The splits overlap if

$$u_i < v_j \wedge u_j < v_i \tag{2.17}$$

otherwise we can impose constraints functionally similar to (2.16) but without the red clearance interval, which is only relevant to incompatible phases. We write

$$\left.\begin{aligned} u_i &= v_j \\ u_j &= v_i \end{aligned}\right\} \forall j \in S_i, i = 1..m \text{ and } x_{ij} = 0 \tag{2.18}$$

$$\left.\begin{aligned} v_i &\leq y_{ij} + u_j \\ v_j &\leq (1 - y_{ij}) + u_i \end{aligned}\right\} \forall j \in S_i, i = 1..m \text{ and } x_{ij} = 1 \tag{2.19}$$

Next, given $x_{ij} = 0$ one must decide on additional constraints with respect to the characteristic of the overlap of the splits. We propose it must hold that

$$u_i = u_j \wedge v_i = v_j \tag{2.20}$$

that can be written

$$\left.\begin{aligned} u_i &= u_j \\ v_i &= v_j \end{aligned}\right\} \forall j \in S_i, i = 1..m \text{ and } x_{ij} = 0 \tag{2.21}$$

Note that (2.21) dominates constraints (2.18). We now examine the impact on h_i due to the priority movement h_j.

- **Free flow:** During free flow in h_j we assume a linear reduction of the saturation flow rate given by

$$s_i' = s_i p_j (1 - x_{ij}) \tag{2.22}$$

where $p_j = 1 - \frac{q_j}{s_j}$ is the probability that the flow in h_j has an available gap to let passenger car units from h_i pass. If $\frac{q_j}{s_j} \geq 1$ then the vehicles of h_j maintain the saturation headway which is insufficient in providing any available gaps. In general, h_i must give way to movements h_j for $j \in S_i$ such that its adjusted saturation flow rate is given by

$$s_i' = s_i \prod_{\forall j \in S_i} p_j(1 - x_{ij}) \tag{2.23}$$

since the $p_j s$ are independent events.

- **Conflicting standing queues:** During the red indication a standing queue in h_j is build up. As the signal turns green the queue is discharged at rate s_j during which there will be no available gaps for the passenger car units of h_i to pass. After the standing queue has been cleared free flow will be present for which the conditions described in previous item apply. Therefore, h_i must account for the required green time to clear the standing queue in h_j. In general, given S_i from previous item h_i must take into account the longest standing queue

$$Q_i = \max\left\{x_{ij}\frac{q_j}{s_j}, \forall j \in S_i\right\} \tag{2.24}$$

Combining (2.25) and (2.24) the reduced flow ratio of h_i is given by

$$\zeta_i' = Q_i + \frac{q_i}{s_i'} \tag{2.25}$$

Unfortunately, $\frac{q_i}{s_i'}$ is highly non-linear since h_i may have to give way to several priority movements simultaneously and due to the resulting reciprocal of s_i'. Nonetheless, a simple linearisation can be achieved, however at the expense of restricting h_i such that it may give way to a single priority movement at a time only. This is incorporated by introducing a constraint similar to (2.9) given by

$$\left.\begin{aligned} f\zeta_i' &\leq g_i && \forall j \in S_i \\ \zeta_i' &= \frac{q_j}{s_j} + \frac{q_i}{s_i(1 - q_j/s_j)} \end{aligned}\right\} \forall S_i \in \mathcal{S} \text{ and } x_{ij} = 0 \tag{2.26}$$

$$\sum_{j \in S_i} x_{ij} \leq 1 \qquad \forall S_i \in \mathcal{S} \tag{2.27}$$

Note that $\zeta_i' \geq \zeta_i$[1], thus (2.26) restricts the effective green ratio of h_i to be larger and therefore dominates (2.9) in the presence of an overlap.

Finally, substituting $u_i = v_i - g_i - z\ell_i$ from (2.8) into all relevant constraints the position of the split is determined solely be v_i. From (2.10,2.11) the bounds on v_i become apparent

$$\left.\begin{aligned} v_i &\geq g_h + z\ell_i \\ v_i &\leq 1 + g_h + z\ell_i \end{aligned}\right\} \forall h_i \in H \tag{2.28}$$

[1]Recall that ζ_i is the flow ratio for some movement $h_i \in H$ in the absence of overlaps.

2.1.3.4 Model summary

Model 2.1.1: Extended model of (IC)

$$1 \leq z\bar{c}$$
$$1 \geq z\underline{c}$$
$$\left.\begin{aligned} z\underline{g}_h &\leq g_i \\ g_i &\geq f\zeta_i \\ v_i &\geq g_i + z\ell_i \\ v_i &\leq 1 + g_i + z\ell_i \end{aligned}\right\} \forall h_i \in H$$
$$\left.\begin{aligned} z\ell^{ij}_{rci} + v_i - (v_j - g_j - z\ell_j) &\leq w_{ij} \\ z\ell^{ji}_{rci} + v_j - (v_i - g_i - z\ell_i) &\leq (1 - w_{ij}) \end{aligned}\right\} \forall (i,j) \in I$$
$$\left.\begin{aligned} v_i - (v_j - g_j - z\ell_j) &\leq y_{ij} \\ v_j - (v_i - g_i - z\ell_i) &\leq (1 - y_{ij}) \end{aligned}\right\}^{1)} \forall j \in S_i, i = 1..m \text{ and } x_{ij} = 1$$
$$\left.\begin{aligned} v_i - g_i - z\ell_i &= v_j - g_j - z\ell_j \\ v_i &= v_j \end{aligned}\right\}^{2)} \forall j \in S_i, i = 1..m \text{ and } x_{ij} = 0$$
$$\left.\begin{aligned} f\zeta_{ij} &\leq g_i \qquad \forall j \in S_i \\ \zeta_{ij} &= \frac{q_j}{s_j} + \frac{q_i}{s_i(1 - q_j/s_j)} \end{aligned}\right\}^{3)} \forall S_i \in \mathcal{S} \text{ and } x_{ij} = 0$$
$$\sum_{j \in S_i} x_{ij} \leq 1 \qquad \forall S_i \in \mathcal{S}$$

The sets of constraints, 1) through 3), can be implemented in `GUROBI` using the method

```
model.addConstrs((conditions) >> (constraint) for a range of values)
```

If `conditions` are true then `constraint` is considered, else it is ignored.

2.1.3.5 Objective function

We investigate the following objectives.

1. Maximization of the intersection capacity factor.
2. Maximization of the sum of effective green ratios.
3. Minimization of the cycle length (maximization of z).
4. Minimization of the delay. Webster's formula [11] describes the average delay (in seconds) per cycle per passenger car unit served by some movement $h_i \in H$ given by

$$d_i(g_i, c) = 0.9 \cdot \left(\frac{c(1 - g_i/c)^2}{2c(1 - q_i/s_i)} + \frac{3600b^2}{2q_i(1 - b)} \right) \tag{2.29}$$

where $b = \dfrac{q_i c}{s_i g_i}$ is the degree of saturation. Note that g_i is an absolute duration rather than a ratio of the cycle length. The first term is due to standing queue

build ups during the effective red indication. The second term is the contribution due to random poisson arrivals investigated by Webster, the main idea being that the random arrivals nearer the saturation level contribute to further delay. Finally, Webster considered a third term meant to decrease the travel time, but was simplified by a multiple factor of 0.9. Given a constant value for g_i (2.29) suggests the cycle length to be minimized.

Given a set $O = \{o_1, ..., o_n\}$ of objectives there are two ways to handle a multi-objective problem

- one defines a lexicographical order on O, say $o_1 < o_2 < ... < o_n$, such that the optimal value of o_i is accepted only if it does not degrade the values of $o_{j=1..i \mid j<i}$,
- one assigns to each objective $o_i \in O$ a weight $w_i \in \mathbb{R}$ from which a single objective is derived by the linear combination $f(O) = \max \sum w_i o_i$.

Some notes on objectives 1) through 4) are as follows.

- Objectives 1) and 2) are positively correlated, however not strictly speaking as it may be possible only to increase the green of certain movements whilst f must remain unchanged due to other movements whose green can not be increased.
- The *critical* cycle length is the required cycle length such that $f = 1$.
- In the following chapters we seek to estimate the optimal cycle length based on simulation. Figures 2.2a and 2.2b illustrates a tentative result from such simulation experiment. We solved model (IC) for pre-fixed cycle lengths $c = \{20, 21, ..., 150\}$. Using a traffic signal simulation (Chapter 3) we estimated the average delay for each cycle length (orange line) given 5 simulation iterations each of duration 12 hours. A minimum at $c \approx 60$ can be concluded.

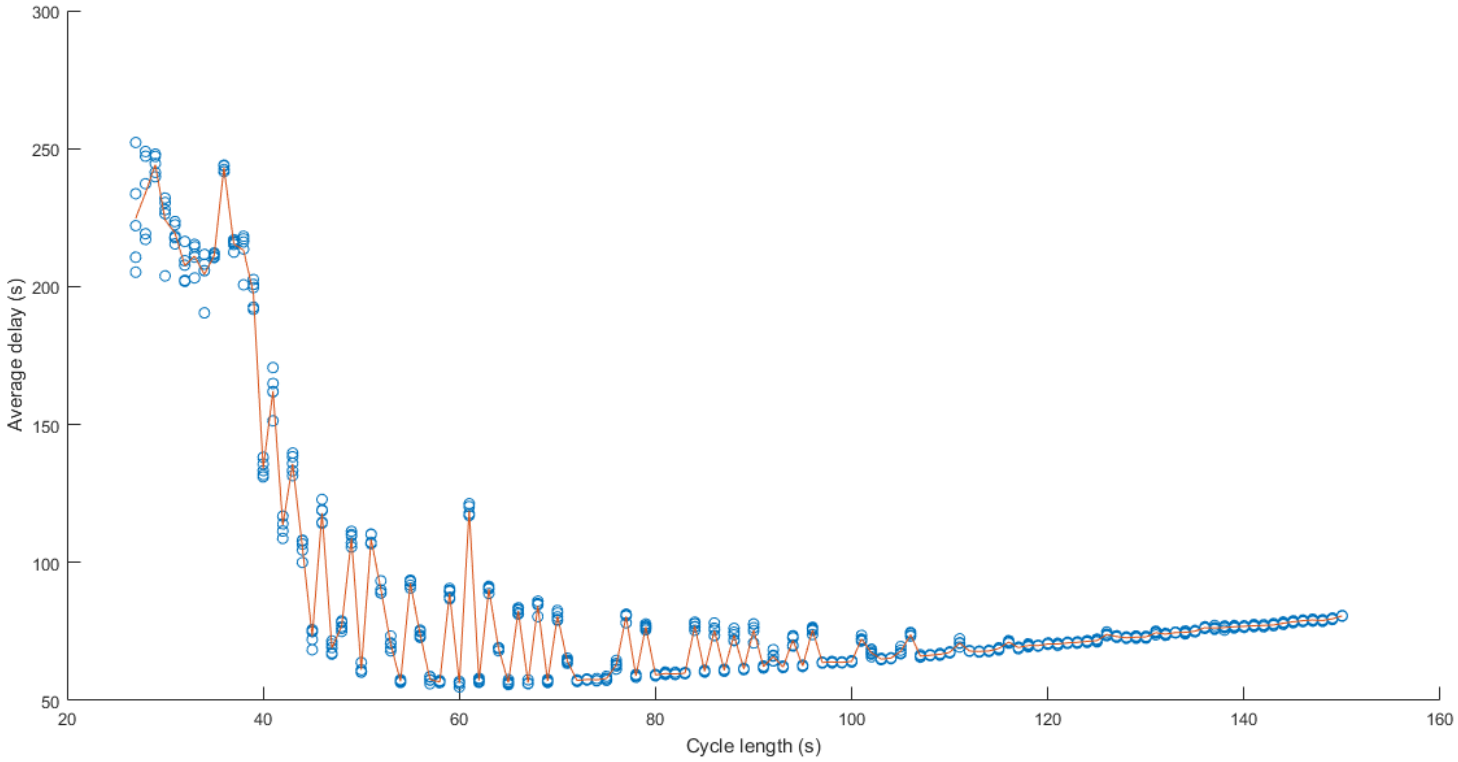
2.1.4 An extended version, (GIv4), of (GIv3)

Once again, the details of the extensions are highlighted in the model description.

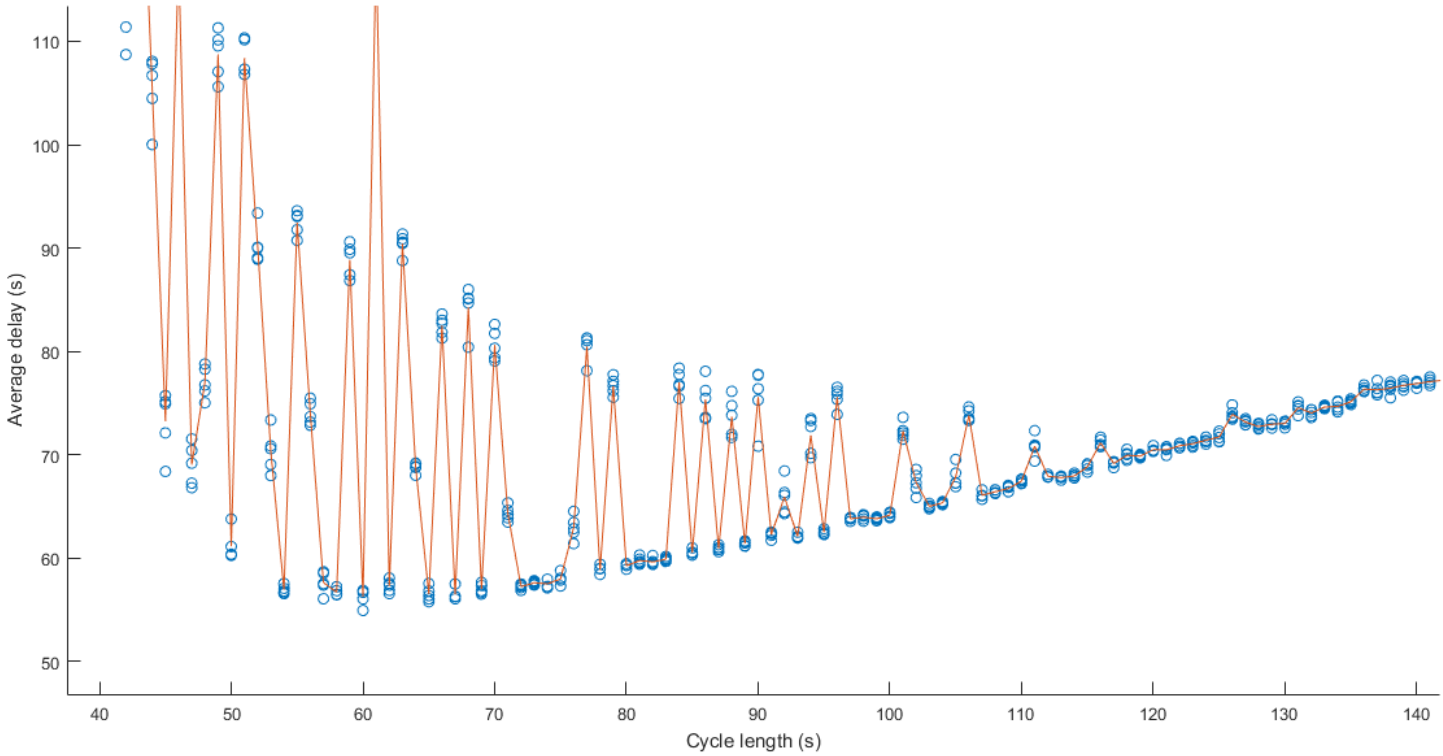
- Given a set of movements H the model (GIv4) considers the set of all phases $\mathcal{K} = \mathcal{P}(H) \setminus \emptyset$. At first sight each phase $K \in \mathcal{K}$ is unconditional in the sense that two movements $h_a, h_b \in K$ may be incompatible, but as pointed out in chapter 1 such a phase is infeasible due to the lack of a priority order. Thus, let $\mathcal{K}_{con} \subseteq \mathcal{K}$ be the set of conditional phases such that no phase contains a pair of incompatible movements. Note that $\mathcal{K}_{con}$ is *not* the phase specification meaning that each phase is not required to be a part of the final phase specification. Similar to (GIv2) we constrain

$$\bigcup_{\forall K \in \mathcal{K}_{con}} K = H \qquad \text{but not} \qquad \bigcap_{\forall K \in \mathcal{K}_{con}} K = \emptyset$$

thus final phase specification may contain multiple phases sharing a common movement. Consequently, depending on the phase scheduling a movement may receive more than one green interval in a cycle. The main task of this model is then to maximize the number of phases and their split durations such that each movement appears in at least one phase while ensuring a feasible phase scheduling.



(A)



(B) Zoom.

FIGURE 2.2

Input parameters

- Let $H = \{1, ..., m\}$ be the set of unique movements of the intersection in question.
- Let $\mathcal{K}_{con} = \{K_1, ..., K_n\}$ be an ordered set of phases such that $K_i \subseteq H$ for $i = 1..n$ and
$$\bigcup_{\forall K \in \mathcal{K}_{con}} K = H$$
- Define the sets $S = \{(I, J) \in \mathcal{K}_{con} \times \mathcal{K}_{con}$ s.t. I, J are semi-compatible$\}$ and $P = \{(I, J) \in \mathcal{K}_{con} \times \mathcal{K}_{con}$ s.t. $I < J \wedge I, J$ are incompatible$\}$.
- Let $c \geq 0$ be the cycle length.
- Let q_h, s_h be the arrival rate and saturation flow rate, respectively, for some movement $h \in H$.
- Let ℓ_{IJ}^{rci} be the red clearance interval for incompatible phases $I, J \in \mathcal{K}_{con}$ when transitioning from I to J. Note that we may have $\ell_{IJ}^{rci} \neq \ell_{JI}^{rci}$.
- Given a movement $h \in H$ the following parameters hold.
  - Let $r_{h,\min}, r_{h,\max}$ be its minimum and maximum effective red.
  - Let $g_{h,\min}, g_{h,\max}$ be its minimum and maximum effective green.
  - Let ℓ_h be its lost time, i.e., the sum of the start-up delay and clearance lost time.
- Given a phase $K \in \mathcal{K}$ the following parameters hold.
  - Let $r_{K,\min} = \max\{r_{h,\min}, \forall h \in K\}$ be its minimum and $r_{K,\max} = \min\{r_{h,\max}, \forall h \in K\}$ be its maximum effective red, respectively.
  - Let $g_{K,\min} = \max\{g_{h,\min}, \forall h \in K\}$ be its minimum and $g_{K,\max} = \min\{g_{h,\max}, \forall h \in K\}$ be its maximum effective green, respectively.
  - Let $\ell_K = \max\{\ell_h, \forall h \in K\}$ be its lost time.
  - Let (as derived in version 1)
$$\zeta_K = \max\{q_i/s_i', \forall i \in K\} \tag{2.30}$$
$$s_i' = s_i \prod_{\forall j \in K_i} (1 - q_j/s_j) \tag{2.31}$$
    be its critical movement flow ratio.

Decision variables

- Let $\mathcal{K}_{con}' \subseteq \mathcal{K}_{con}$ be the set of phase specifications.
- Let $p_K \in \{0, 1\}$ be a binary variable defined for each phase $K \in \mathcal{K}_{con}$ such that $p_K = 0$ if $K \in \mathcal{K}_{con}'$ and $p_K = 1$ otherwise.
- Let $f \geq 0$ be the *intersection capacity factor*. This is a maximum common multiple factor with respect to all phases.
- Let $w_{IJ} \in \{0, 1\}$ be a binary variable defined for each $(I, J) \in P$ such that $w_{IJ} = 0$ if I precedes J, and $w_{IJ} = 1$ otherwise, i.e., J precedes I.

- Let $x_{IJ} \in \{0,1\}$ be a binary variable defined for each $(I,J) \in S$ such that $x_{IJ} = 0$ if the splits of I and J overlap and $x_{IJ} = 1$ otherwise.
- Let $y_{IJ} \in \{0,1\}$ be a binary variable defined for each $(I,J) \in S$ given $x_{IJ} = 0$ such that $y_{IJ} = 0$ if the split of I precedes that of J and $y_{IJ} = 1$ otherwise.
- Given a phase $K \in \mathcal{K}_{con}$ the following variables hold.
  - Let u_K, v_K be the starting time of G and ending time of Y, respectively, such that $v_K - u_K$ is the split of K.
  - Let g_K be the effective green.

Constraint formulations

All constraints from model version 1 carry over to this model, but additional constraints must be considered.

We must have

$$\bigcup_{\forall K \in \mathcal{K}'_{con}} K = H \tag{2.32}$$

(2.33)

Each movement $h \in H$ must appear in at least one of the phases in $\mathcal{K}'_{con}$

??? (2.34)

The splits $[u_I, v_I]$ and $[u_J, v_J]$ of a pair of semi-compatible phases $(I,J) \in S$ are permitted to overlap *but it is not required*. The splits overlap if

$$u_I < v_J \wedge u_J < v_I \tag{2.35}$$

otherwise we can impose constraints functionally similar to (**??**) but without the red clearance interval, which is only relevant to incompatible phases. We write

$$\left.\begin{array}{ll} u_I - v_J + \epsilon & \leq M x_{IJ} \\ u_J - v_I + \epsilon & \leq M x_{IJ} \\ v_I - u_J & \leq M(1 - x_{IJ}) + M y_{IJ} \\ v_J - u_I & \leq M(1 - x_{IJ}) + c + M y_{IJ} \\ v_J - u_I & \leq M(1 - x_{IJ}) + M(1 - y_{IJ}) \\ v_I - u_J & \leq M(1 - x_{IJ}) + c + M(1 - y_{IJ}) \end{array}\right\} \forall (I,J) \in S \tag{2.36}$$

where M is a large constant and $\epsilon > 0$ is introduced to handle the apparent strict inequalities. In model version 2 we constrained the split of a non-priority movement to be included in that of the priority one. However, given two phases $I, J \in \mathcal{K}_{con}$, I may contain a priority movement with respect to some movement in J and/or vice versa. Therefore, we must constrain the splits of I and J be equal and overlap if $x_{IJ} = 0$

$$u_I = u_J \wedge v_I = v_J \tag{2.37}$$

that translates into

??? (2.38)

Finally, we need to consider the semi-compatibilities among $I \times J$ for $I, J \in \mathcal{K}_{con}$ with respect to reduced saturation flow rates.

Objective function

The objective function remains almost unchanged with respect to the one in version 1 (and version 2). We must consider the maximization of the number of phases in the phase specification

$$\max \sum_{\forall K \in \mathcal{K}_{con}} p_K \tag{2.39}$$

Remark

Given a large number of movements the set $\mathcal{K}_{con}$ may render the model computationally infeasible. Alternatively, one may introduce the set $\mathcal{K}_{fea} \subseteq \mathcal{K}_{con}$ of phases chosen for the sake of computational feasibility.

Model summary

Model 2.1.2: Model (GIv4)

2.1.5 Implementation details

For any intersection it is sufficient specifying only the compatibilities between each pair of movements, as one may then derive necessary relationships among movements composing individual phases (with respect to phase specifications), and among each pair of phases (with respect to phase orderings).

2.1.6 Examples

We consider a four-way intersection with exclusive left and right turn lanes and pedestrian lane for each approach whose movements are numbered as shown in Figure 2.3. The cycle length is a parameter, set to $c = 60$. The example is solved by means of the above model versions 1 and 2.

- Version 1 - we apply the set of phase specifications from Figure 2.4 (non-singleton).
- Version 2 - we consider each movement as a single phase (singleton).

Figure 2.5 and 2.6 illustrate the solutions.
Model version 2 is superior to version 1 in terms of both objectives.

2.2 Fixed-time coordinated signals

2.2.1 MAXBAND

MAXBAND [6] aims to facilitate the situation in which vehicles are able to travel unimpeded by traffic signals along a major arterial road containing multiple intersections due to a "green wave". As already mentioned coordinated control generally

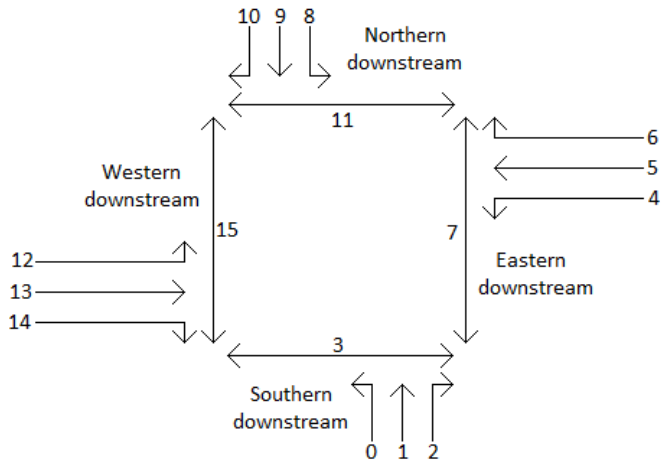


FIGURE 2.3: Movement numbering in a four-way intersection with exclusive left and right turn lanes. Note, each downstream is a single lane e.g. the northern downstream collects incoming traffic from movement 1, 6 and 12. In this example we assume an unique traffic signal assigned to each movement, i.e., the number of lanes equal the number of movements.

deals with offsets and green bandwidths that are both encapsulated in a time-space diagram.

2.2.2 Coupling with model (IC)

We propose a model that combines the essence of model (IC) and MAXBAND in order to determine simultaneously the phase specification, phase schedule, splits, cycle length and offsets.

2.2.2.1 Preliminaries

Demand estimation in coordination

In any coordinated system a movement demand of some intersection may be dynamic as it depends on the employed signal settings of adjacent intersections. Figure 2.7 illustrates two intersections whose input demands q_i are given from the exterior surroundings. Given turn rates can be used to determine individual movement demands. The demands of movements marked ? can be dealt with mainly in two ways.

- In a straight forward approach the unknown demands can be found by solving a network flow problem given exterior arterial demands q_i and turn ratios. An apparent disadvantage is the fact that one assumes all-green signal settings thus allowing each movement flow to simply add up forming other movement flows.
- Given actual signal settings the movement demands marked ? depend on the splits of relevant adjacent movements. Since a split can be at most the cycle length the process in previous item always overestimates the unknown movement demands.

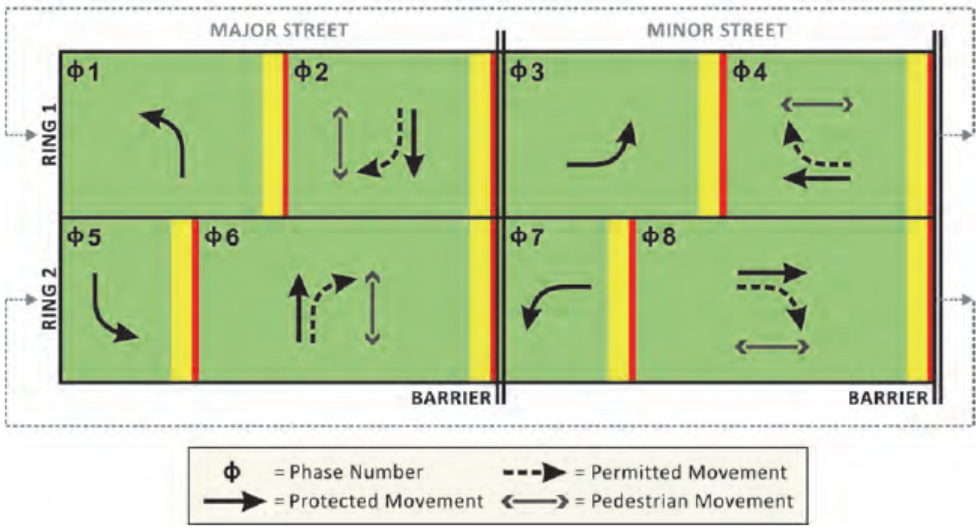


FIGURE 2.4: **ADD CAPTION**

2.2.2.2 Input parameters

- Let $N = \{1, ..., n\}$ be the set of intersections. They are to be imagined with intersection 1 being the left-most and n the right-most that we refer to as *boarder* intersections.
- Let $\mathcal{H} = \{H_1, ..., H_n\}$ be the set of sets of movements for each intersection such that $H_i = \{h_1, h_2, ...\}$ for each $i \in N$.
- Define a global partial order (order of priority) on each H_i such that $h_a, h_b \in H$
  - are *incomparable* if they are compatible,
  - have equivalent priority, $h_a \equiv h_b$, if they are incompatible.
- Define the sets
  - $\mathcal{I} = \{I_1, ..., I_n\}$ such that $I_i = \{(a, b) \in \{1..|H_i|\} \times \{1..|H_i|\} \mid h_a \equiv h_b\}$ for each $i \in N$.
  - $\mathcal{D} = \{D_1, ..., D_n\}$ such that $D_i \subseteq H_i$ is the set of movements of intersection $i \in N$ that has a dynamic demand. These are the movements that approach the intersection from any adjacent interior intersections.
  - $\overleftarrow{\mathcal{H}} = \{\overleftarrow{H}_1, ..., \overleftarrow{H}_{n-1}\}$ such that $\overleftarrow{H}_i \subseteq H_i$ is the set of movements that supply demands to the dynamic movements of intersection $i + 1$.
  - $\overrightarrow{\mathcal{H}} = \{\overrightarrow{H}_2, ..., \overrightarrow{H}_n\}$ such that $\overrightarrow{H}_i \subseteq H_i$ is the set of movements that supply demands to the dynamic movements of intersection $i - 1$.
  - $\mathcal{S} = \{S_1, ..., S_n\}$ such that $S_i \subseteq H_i$ is the set of movements of intersection $i \in N$ that has a static demand. These are the movements that approach the intersection from any adjacent exterior intersections.
  - $\mathcal{L} = \{L_1, ..., L_n\}$ (major arterial left turn ratios) such that $L_i = \{[0, 1], [0, 1]\}$ for $i = 2..n - 1$ and $L_1, L_n \in [0, 1]$, where L_{i1} (L_{i2}) for $i = 2..n - 1$ is related to the outbound (inbound) direction.

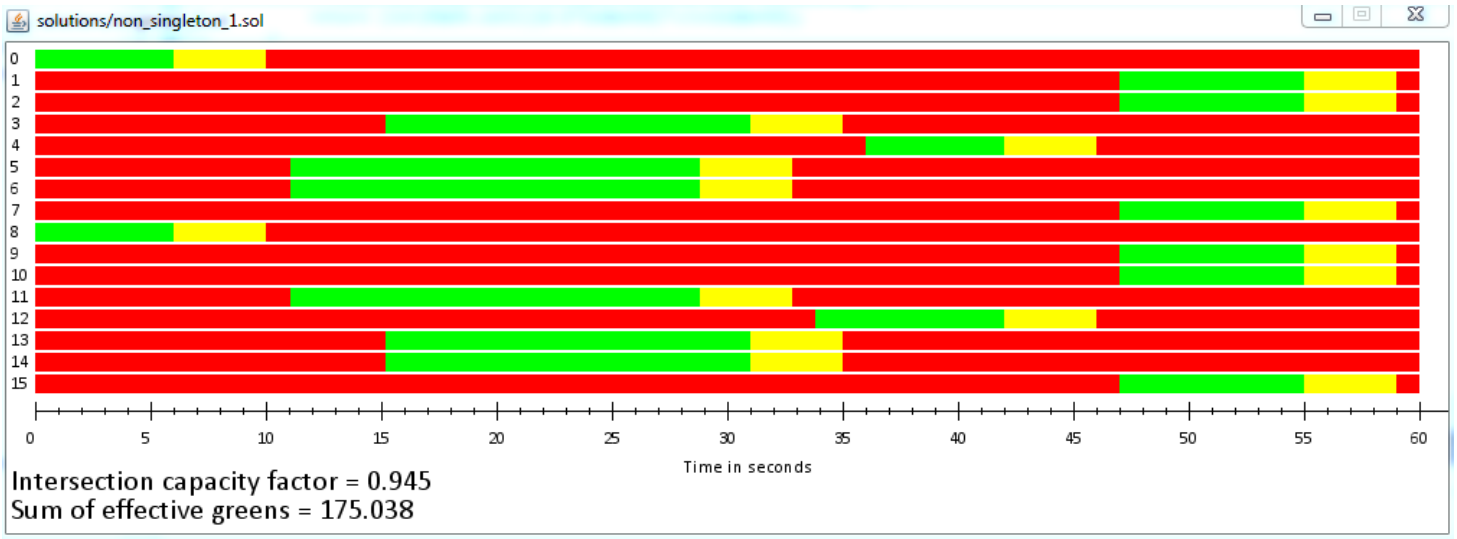


FIGURE 2.5: Time signal plan using predetermined phase specifications. Note, the plot illustrates the display sequence for each individual movement although several movements compose a common phase.

  - $\mathcal{R} = \{R_1, ..., R_n\}$ (major arterial right turn ratios) such that $R_i = \{[0,1],[0,1]\}$ for $i = 2..n-1$ and $R_1, R_n \in [0,1]$, where R_{i1} (R_{i2}) for $i = 2..n-1$ is related to the outbound (inbound) direction.

- For each road segment (section between intersections) define the allowed speed interval.
- Let $\ell_{rci}^{ij} \geq 0$ (seconds) be the red clearance interval for $(i,j) \in I$ when transitioning from h_i to h_j. Note that we may have $\ell_{rci}^{ij} \neq \ell_{rci}^{ji}$.
- For each $h_i \in H$ we define (all in seconds)
  - $\underline{g}_i$ to be its minimum effective green.
  - ℓ_i to be its lost time, i.e., the sum of start-up delay and clearance lost time. Note, in contrast to the definition in Chapter 1 the ℓ_i in this model does *not* include the red clearance interval, ℓ_{rci}^{ij}. Instead, it is specified independently, see above.
- Let $q_i, s_i \geq 0$ (both passenger car units per hour) be the arrival rate and saturation flow rate, respectively, for some movement $h_i \in H$.
- Let $\zeta_i = \dfrac{q_i}{s_i}$ be the flow ratio that restricts the required effective green ratio, g_i (to be defined), of some movement $h_i \in H$. The flow ratio is justified as follows. During the effective red ratio, $1 - g_i$, the expected standing queue build up per cycle is given by $q_i(1 - g_i)$. During g_i the passenger car units are discharged at rate $s_i - q_i$ due to additional vehicle arrivals while clearing the standing queue such that movement h_i may serve $g_i(s_i - q_i)$ passenger car units per cycle. Thus, it must hold that

$$(s_i - q_i)g_i = q_i(1 - g_i) \Leftrightarrow \tag{2.40}$$

$$g_i = \frac{q_i}{s_i} \tag{2.41}$$

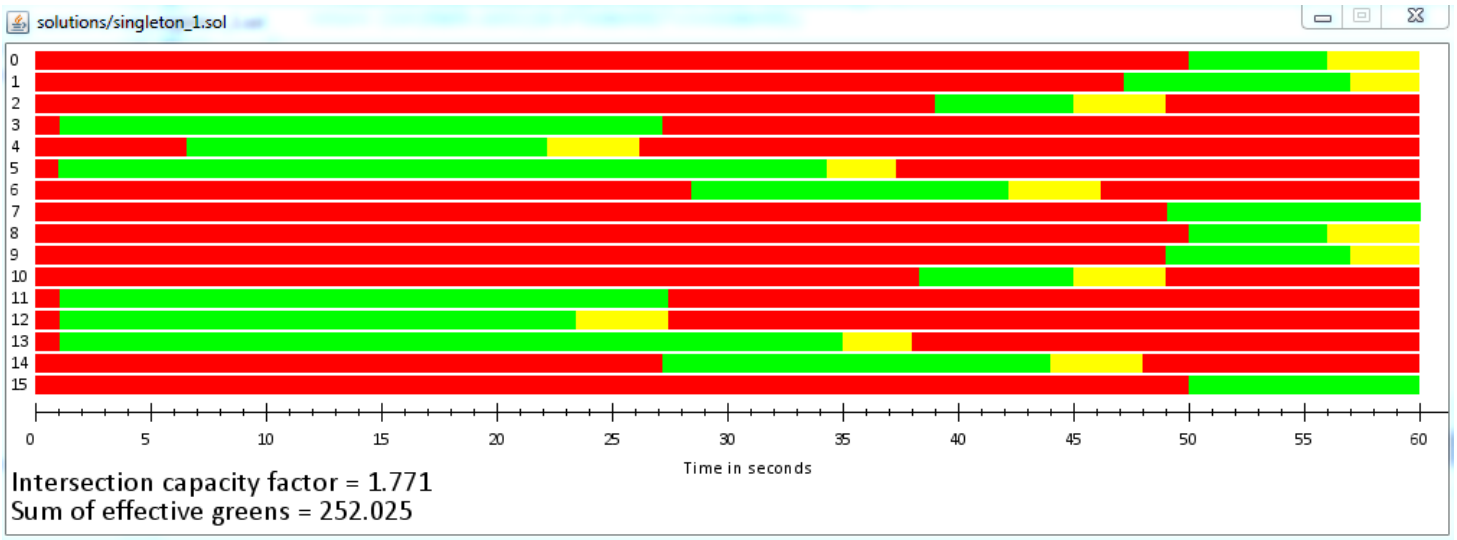


FIGURE 2.6: ADD CAPTION

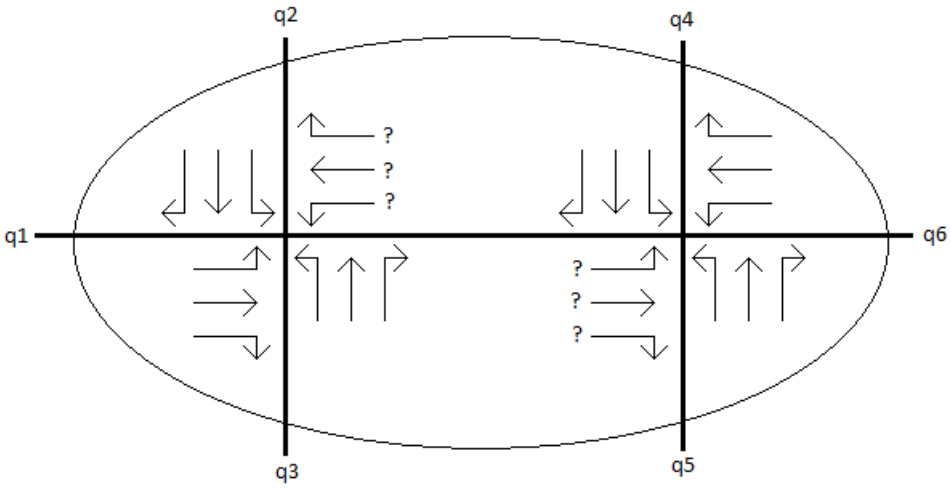


FIGURE 2.7

A traffic signal is not operating effectively when it displays green for a movement with free flow compared to when discharging a standing queue. This is due to potential standing queue build ups in every other movement, hence the equality sign in (2.41). Nevertheless, one should always allow a larger green ratio that does not affect the queue handling of every other movement, hence one should consider (2.41) be

$$g_i \geq \frac{q_i}{s_i} \tag{2.42}$$

In the constraint section we will introduce a constraint with respect to (2.42).

- The model (IC) requires the demand specified for each movement. However, given successive signals along a corridor the demand of certain movements depends on the signal settings. For instance, in Figure 2.7 the demand of movement 13 of intersection B depends on the green ratios of movements 2,8 and 13

of intersection A.

$$q_{B,13} = (1 - R_{B,W} - L_{B,W})(g_{A,2}q_{A,2} + g_{A,8}q_{A,8} + g_{A,13}q_{A,13} \tag{2.43}$$

However, each movement demand is multiplied by the intersection capacity factor, and in the context of semi-compatible movements one divides by the demand. Thus $q_{B,13}$ imposes several non linearities. In order to bypass those issues we can disregard the intersection capacity factor and instead maximize only the sum of effective green ratios. In addition, we may simply consider each pair of semi-compatible movements rather to be incompatible.

2.2.2.3 Decision variables

- For each movement $h_i \in H$ we define
  - g_i to be its effective green ratio.
  - u_i to be its starting time of G and v_i its ending time of Y such that $v_i - u_i$ is the split ratio of h_i that accounts for both g_i and ℓ_i.
- Let $f \geq 0$ be the intersection capacity factor. This is a maximum common multiple factor with respect to all movements.
- Let z be the reciprocal cycle length $\frac{1}{c}$ such that $\frac{1}{\bar{c}} \leq z \leq \frac{1}{\underline{c}}$ where $\underline{c}$ is the lower bound and $\bar{c}$ the upper bound (both seconds) on the cycle length c.
- Let $w_{ij} \in \{0, 1\}$ be a binary variable defined for each $(i, j) \in I$ such that $w_{ij} = 0$ if the split of h_i precedes that of h_j and $w_{ij} = 1$ otherwise, i.e., the split of h_j precedes that of h_i.
- Let b ($\bar{b}$) be the green bandwidth in the outbound (inbound) direction.
- Let π_i ($\bar{\pi}_i$) be the time from the right side of the red at intersection i to the left side of the green bandwidth in the outbound (inbound) direction.
- Let $t(h, i)$ ($\bar{t}(h, i)$) be the travel time from intersection h to i in the outbound (i to h in the inbound) direction.

2.2.2.4 Constraints

$\pi \geq$ sum of cycle proportion demands due to through vehicles of minor turning lanes from previous inter (2.44)

Dynamic demands

Given an outbound dynamic left, right and through movement, say $h_1, h_2, h_3 \in D_i$, their demands are given by

$$q^i_{h_1} = L_{i1} \sum_{j \in \overleftarrow{H}_{i-1}} q^{i-1}_j g^{i-1}_j \tag{2.45}$$

$$q^i_{h_2} = R_{i1} \sum_{j \in \overleftarrow{H}_{i-1}} q^{i-1}_j g^{i-1}_j \tag{2.46}$$

$$q^i_{h_3} = (1 - L_{i1} - R_{i1}) \sum_{j \in \overleftarrow{H}_{i-1}} q^{i-1}_j g^{i-1}_j \tag{2.47}$$

2.2.2.5 Objective function

- Maximize $\sum f_i$ and examine from the solution the green bandwidths.
- Maximize the sum of outbound and inbound bandwidths, and then $\sum f_i$ such that an improvement in the latter is accepted only if it does not degrade the value of the former objective.

2.2.2.6 MILP model summary

Given a set of pre-optimized fixed time isolated signals they must share a common cycle length in order to be coordinated.

Consider the above nameless method being applied individually to n consecutive intersections $S_1, \ldots, S_n$. The MAXBAND method [6] determines the optimal offsets based on the determined parameters for each S_i, however, it assumes a common cycle time among all intersections that is unlikely to happen due to each intersection being treated separately. A proposed solution is to apply the nameless method to each intersection followed by another application in which the cycle time is constrained to be equal to the maximum cycle time found among all intersections during the initial application. This is reasonable since an increased cycle time has more effective green time available for allocation among the phases.

Consider two intersection like that in Figure 2.3. The relevant movements in intersection I1 are 2, 13 and 8 as they lead traffic to the other intersection. Likewise, the relevant movements in intersection I2 are 0, 5 and 10. The traffic of every other movement do not affect the coordination explicitly. Of course, it does affect in some way in the sense of that they must be considered also in the traffic signal plan. Overall, we must enforce some relationship between movements 2, 13 and 8 from west to east, and movements 0, 5, and 10 from east to west.

Despite the integer constraint between intersection cycles it may be reasonable to coordinate in a single direction only, for instance morning and afternoon rush hour depending on if people mainly leave or enter a city at those times of the day.

Changing average flow velocities is reasonable between successive intersections as it may turn out to be beneficial to either go a bit faster or slower due to geometric constraints.

If $r_i = \bar{r}_i$ and $\delta_i = 0$ for $i = 1..n$ then $\phi, \bar{\phi} \in \left\{ \frac{\mathbb{N}}{2} \right\}$ for $i = 1..n$.

Does it make sense that the distance between two intersections is different between the outbound and inbound direction? Yes, it does. It may occur if the road segment between the intersection curve.

An important assumption to coordinated successive arterial intersections is that the contributing vehicles due to turns in all intersections are outnumbered by the through vehicles. If not, then one must optimize the entire network.

It is assumed that each platoon upon reaching the next intersection is not hindered by stationary vehicles at that intersection.

2.2.2.7 Left turn phase sequence

In considering an intersection with an arterial direction whose demands are significantly greater than minor approaches the left turning movement is the main issue for the opposing through movement. Left turn phasing is the task of determining the phase sequence between left turns and the opposing through movements. Left turn phasing can have significant impact on the green bandwidth.

2.3 Traffic-responsive isolated signal

The *vehicle-interval method* [8] utilizes real-time measurements from induction loops. It is applicable to an intersection whose movements may be served using only two phases, possibly a three-way (T-shaped) intersection. As the name suggests it does not consider pedestrian and bicycle movements. Each upstream induction loop is located some distance behind the intersection. Minimum green durations (possibly different) are assigned to both phases. Consider the beginning of one of the two phases at an arbitrary point in time. The method asks whether a vehicle passed the induction loop during the minimum green duration. If so, the r.o.w. is extended to a degree that allows the vehicle to pass the intersection based on the speed limit. In the extended r.o.w. the method keeps asking if additional vehicles pass the induction loop, whose presence similarly extends the r.o.w. The process goes on until some maximum green time has been reached, at which the other phase is employed. The method may be deemed a reactive strategy, as it constantly requires yes/no answers only.

2.4 Traffic-responsive coordinated signals

Chapter 3

Traffic signal simulation

3.1 Motivation

The evaluation strategy of a traffic control system, as described in chapter 1, concerns the task of estimating performance measures based on simulation studies. In particular, this chapter addresses the usage of microscopic simulations.

The general concept of a microscopic traffic signal simulation is that each traffic user is treated individually as separate entities that interact and behave accordingly to its surroundings. A common application is by means of *car-following models* in which each vehicle adjusts its velocity and acceleration with respect to the one in front of it.

3.2 Review of car-following models

3.3 A microscopic traffic signal simulation

We present and discuss a simple microscopic traffic signal simulation. It is made in the Java programming language.

3.3.1 The simulation loop

In general, a simulation loop controls the overall flow of any computer/video game and/or simulation.

- A *tick* refers to a single action in which the logic is updated once with respect to some time step, in particular the position, velocity and acceleration of each vehicle.
- The *rendering* refers to the drawing of every object (vehicles, traffic signals, roads etc.) onto the screen.

A common game loop example design of highest simplicity employing the above concepts is given in **Algorithm** 1.

Algorithm 1

```
1: procedure RUNSIMULATION
2:     while isRunning do
3:         update game logic (tick)
4:         draw to screen (render)
5:     end while
6: end procedure
```

A crucial task in developing a game loop is to ensure a smooth real time simulation flow independent of the hardware of the device on which it is being run. In fact, the game loop 1 exhibits certain flaws as it does not keep track of nor does it measure the time passed, therefore a fast computer would run through the loop at a much greater rate than what would be a conceivable simulation. As already explained the ticks control the game logic, thus they are the ones on which to impose time constraints in order to force a specific number of ticks per second (TPS). In contrast, the rendering frequency, also known as the number of frames per second (FPS), may be as high as possible only limited by the capabilities of the computer. One may argue, however, that rendering a frame should take place only between two ticks to avoid unnecessary utilization of computing resources as the logic of the simulation effectively does not change otherwise.

The above concepts are implemented in Java as follows.

```
nspt = 1_000_000_000.0 / tps;
delta = 0;
lastTime = System.nanoTime();
while (true) {
    now = System.nanoTime();
    delta += (now - lastTime) / nspt;
    lastTime = now;
    while (delta >= 1) {
        tick();
        delta--;
    }
    render();
}
```

The parameters are as follows.

- `TICKS1` is the number of ticks per second.
- `nspt` is the time interval between two consecutive ticks.
- `delta` accumulates the current time taken as a fraction of `npts`.
- `lastTime` and `now` are timestamps that are used in order to measure the real time simulation advancement.

Figure 1 illustrates the game loop flow details starting at the `while` initialization and onwards 1 second. During the first `nspt` period there is no tick. This is reasonable as the initial game logic should be rendered prior to any tick updates. The following `nspt` periods then always first carries out a tick followed by as many renderings as manageable for the computer during the remainder of `nspt`.

DESCRIBE THE DIFFERENCE BETWEEN SIMULATION TIME AND REAL TIME. The execution time of the simulation is what we can speed up, the simulation time IS the same as the real time.

In addition to `TICKS1` there are two crucial parameters, however not apparent in the game loop, associated with ensuring a smooth real time simulation. In addition, they also ensure a 1:1 scale compared to the real world.

- `TICKS2` is the simulation time advancement per tick expressed as a fraction of 1 second. For instance, if `TICKS1 = TICKS2 = 1` then the simulation runs in real time since we are executing 1 tick per second and each tick is updating the game logic during a time interval of 1/1 of a second, however, the resulting

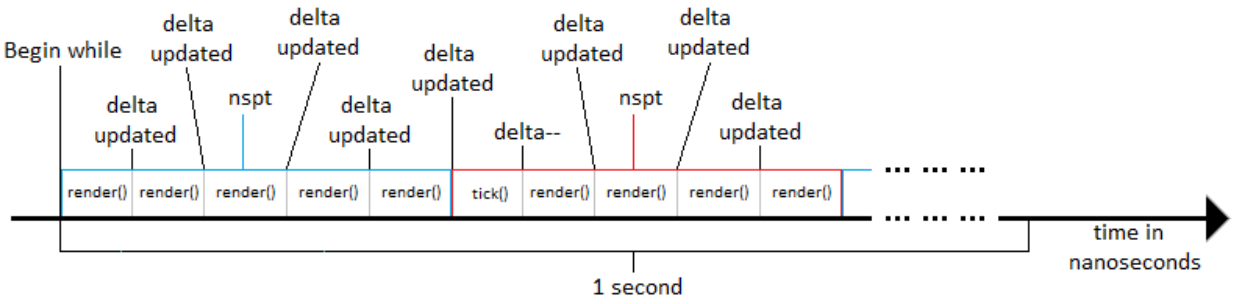


FIGURE 3.1: The game loop flow between ticks and rendering for `TICKS1 = TICKS2 = 60`. Each time interval `nspt` contains a single tick and 4 renderings (in practice it may vary from 3 to 5). Each tick and render block has a duration of `(now - lastTime) / nspt`

simulation is very jumpy. Similarly, if `TICKS1 = TICKS2 = 60` then the simulation also runs in real time since we are executing 60 ticks per second and each tick is updating the game logic during a time interval of $1/60$ of a second. In the latter case, the simulation runs much smoother.

- `PPM` $\in \mathbb{R}$ (pixels per meter) describes the equivalent distance of 1 meter in terms of screen pixels. It serves as a zooming parameter. Given `PPM = 2` and `TAPT` $= \frac{1}{60}$ then the conversion from velocity $13.88\frac{m}{s}$ $\left(\approx 50\frac{km}{h}\right)$ to unit $\frac{pixels}{tick}$ is given by

$$13.88\frac{m}{s} \cdot 2\frac{pixels}{m} \cdot \frac{1}{60}\frac{s}{tick} \approx 0.463\frac{pixels}{tick} \tag{3.1}$$

  thus it becomes evident that moving 0.463 pixels 60 times during 1 second gives rise to smooth movements as stated above, compared to the jumpy situation in which one would move $13.88\frac{m}{s} \cdot 2\frac{pixels}{m} \cdot 1\frac{s}{tick} = 27.76\frac{pixels}{tick}$. Consequently, one should fix `TAPT` $= \frac{1}{60}$[1].

The most useful relationship between `tps` and `TAPT` is the opportunity to speed up the simulation in-time with respect to real time. For instance, to speed up by a factor of 10 one would set `tps` $= 600$.

The game loop does not impose any explicit restrictions on the FPS other than that the specified TPS is prioritized, i.e., if TPS is very large then the FPS considerably drops. However, it is worth mentioning that for small values of `TICKS1` one may replace

```
while (delta >= 1) {
    tick();
    ticks++;
    delta--;
}
render();
frames++;
```

with

```
if (delta >= 1) {
    tick();
    ticks++;
    render();
    frames++;
    delta--;
}
```

in order to force TPS = FPS as to minimize unnecessary rendering.

[1] At present day, the majority of movies and video games are captured at 24 FPS, however, they generally aim towards achieving 60 FPS at which most people perceive smoother movements.

The following is the rendering method. The Java programming language offers the class `BufferStrategy` from which one creates a buffer strategy. Essentially, the buffer strategy preloads rendered images in advance prior to be displayed. This ensures that the next image need not be created when requested. The integer argument (in this case 3, also known as triple buffering) determines the number of preloaded images.

```
1  private void render() {
2      BufferStrategy bs = this.getBufferStrategy();
3      if (bs == null) {
4          this.createBufferStrategy(3);
5          return;
6      }
7      Graphics g = null;
8      try {
9          g = bs.getDrawGraphics();
10
11         ////////////////////////////////
12         /////// Do all rendering ///////
13         ////////////////////////////////
14
15     }
16     finally {
17         g.dispose();
18     }
19     bs.show();
20 }
```

The simulation may run freely without any restrictions on the number of ticks per second. In addition the rendering may be ignored enabling a large number of ticks per seconds. This is helpful especially when assessing a traffic signal schedule and one needs to simulate the scenario spanning, say, 24 hours over multiple iterations (due to randomness and we then take the average).

3.3.2 Dynamic vehicle interactions

The behaviour of a trailing vehicle depends on the characteristic parameters of itself and those of the leading vehicle. A vehicle has the following parameters

- `maxAccel` is the maximum acceleration in any given situation,
- `brakePace` is the desired deceleration upon braking given prevailing conditions,
- `maxBrake` is the maximum brake capability in any given situation,
- `reactionTime` is the elapsed time between the point in time at which the leading vehicle decides to accelerate or brake, and the point in time at which this vehicle responds accordingly,
- `length` is the physical length,
- `leftTurningSpeed` (`rightTurningSpeed`) is the safe velocity at which the vehicle performs a left (right) turn. Reasonable values of the former (latter) are near 60% (40%) of the speed limit,
- `minHW` is the minimum tip-to-tail headway

Real life drivers mostly are clueless about the characteristics of the leading vehicle that may cause collisions due to misjudged appropriate tip-to-tail distances. This model is meant to describe accident-free flows as collisions are almost always caused by abnormal driver behaviours that we do not consider [10].

We consider a type of intelligent driver model in which each vehicle always strives to maintain some *minimum safety headway* (MSH) based on the above parameters.

Definition 1. *Let d be the reaction time of the trailing vehicle. Consider the following two points in time.*

- *At $t = 0$ the leading vehicle starts to brake at its desired brake pace.*
- *At $t = d$ the trailing vehicle responds accordingly and starts to brake at its desired brake pace, assuming its acceleration during the interval $t = [0, d]$ is constant due to its reaction time.*

The MSH is the required tip-to-tail distance at $t = 0$ such that at $t = d$ the corresponding tip-to-tail distance allows the trailing vehicle to come to a full stop w.r.t. its desired brake pace.

The MSH can be derived as follows. Consider two vehicles. At $t = d$ the positions of the tip of the trailing vehicle and the tail of the leading one are given by, respectively

$$s_{trail}(t = d) = \frac{a_{trail}d^2}{2} + v_{trail}(t = 0)d + s_{trail}(t = 0) \tag{3.2}$$

$$s_{lead}(t = d) = \frac{b_{lead}d^2}{2} + v_{lead}(t = 0)d + s_{lead}(t = 0) \tag{3.3}$$

The tip-to-tail distance at $t = d$ is given by

$$s_{lead}(t = d) - s_{trail}(t = d) \tag{3.4}$$

which, according to Definition 1, must be the exact distance required such that the trailing vehicle may come to a full stop w.r.t. its desired brake pace. This brake distance is given by

$$\frac{v_{lead}(t = d)^2 - v_{trail}(t = d)^2}{2b_{trail}} \tag{3.5}$$

where

$$v_{lead}(t = d) = b_{lead}d + v_{lead}(t = 0) \tag{3.6}$$

$$v_{trail}(t = d) = a_{trail}d + v_{trail}(t = 0) \tag{3.7}$$

Therefore we must have

$$b_{trail} = \frac{v_{lead}(t = d)^2 - v_{trail}(t = d)^2}{2(s_{lead}(t = d) - s_{trail}(t = d))} \tag{3.8}$$

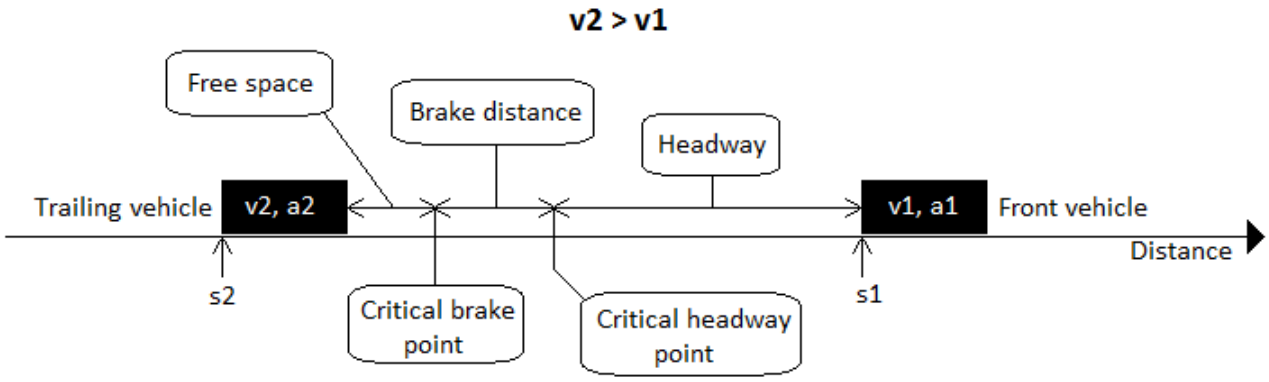


FIGURE 3.2: The interval between each pair of vehicles is divided into three regions.

Finally, we substitute (3.2,3.3) into (3.8 and isolate $s_{lead}(t = 0) - s_{trail}(t = 0)$ that corresponds to the MSH

$$MSH = \frac{(b_{lead}d + v_{lead})^2 - (a_{trail}d + v_{trail})^2}{2b_{trail}} + \left(\frac{a_{trail}d^2}{2} + v_{trail}d\right) - \left(\frac{b_{lead}d^2}{2} + v_{lead}d\right) \tag{3.9}$$

Let h denote the actual tip-to-tail headway. Setting $MSH = h$ we can solve for a_{trail} that is the acceleration function of the trailing vehicle.

$$a_{trail} = \left(b_{trail}d - 2v_{trail} + \sqrt{\sum \alpha_i}\right)(2d)^{-1} \tag{3.10}$$

$$\alpha_1 = 4d^2(b_{lead} - {}^{b_{trail}}/_2)^2 \tag{3.11}$$

$$\alpha_2 = d((4v_{trail} - 8v_{lead})b_{trail} + 8v_{lead}b_{lead}) \tag{3.12}$$

$$\alpha_3 = -8b_{trail}h \tag{3.13}$$

$$\alpha_4 = 4v_{lead}^2 \tag{3.14}$$

If $h > MSH$ then $a_{trail} > 0$ but the velocity is constrained by some speed limit and we must have $a_{trail} \leq$ `maxAccel`. In contrast, the situation in which $a_{trail} < 0$ is limited by `maxBrake`.

he simulation considers three regions encapsulated between two critical points and the tip and tail of the trailing and leading vehicle, respectively. This is illustrated in Figure 3.2. Figure 3.2 illustrates The headway region is always greater than 0 due to a forced minimum headway. The brake region is present only given a difference in velocity. The free space region may occur whether or not the brake region is present.

Upon braking the deceleration is recomputed every tick. This make vehicles brake at non constant deceleration.

- Low reaction time and low brake pace: cautious drivers. Requires large headway.
- High reaction time and high brake pace: aggressive drivers. Requires small headway.
- The last two options 1) low reaction time and high brake pace, 2) high reaction time and low brake pace, require medium headways.

The *minimum safe headway*

3.3.3 Stability analysis

- Explain the role of `EPS`...
- The value of `TAPT` imposes restrictions on `EPS`.
- The value of `maxAccel` must not exceed a certain value (to be explained) in order to ensure stability.

3.3.4 Vehicle spawning

3.3.4.1 Spawning formula

At each tick it is determined if a vehicle may spawn with respect to the arrival rate. There are $3600 \cdot \frac{1}{TAPT}$ ticks per hour, thus we expect $\frac{q \cdot TAPT}{3600}$ spawning pcus per tick. Therefore the arrival rate is respected if, for every tick, we consider

$$p \leq \frac{q \cdot TAPT}{3600} \qquad p \in [0..1] \tag{3.15}$$

As an example, consider $q = 3600 \frac{pcus}{h}$ and $TAPT = \frac{1}{60} \frac{s}{tick}$ then $p \leq \frac{1}{60} \frac{pcus}{t}$. Assuming 60 ticks per second (real time due to $TAPT = \frac{1}{60} \frac{s}{tick}$) we expect 1 pcu per second which is equivalent to q. Alternatively, assuming, say 180 ticks per second, we expect 3 pcus per second, however, the simulation runs 3 times faster than real time, thus the arrival rate is scaled up correspondingly.

3.3.4.2 Spawning variable specification

We assume $a_{trail} = 0$ of a spawning vehicle. Let h be the actual tip-to-tail distance, then the safe spawning velocity is given by

$$v_{trail} = db_{trail} + \sqrt{d^2 b_{trail}^2 + (-d^2 b_{lead} - 2v_{lead}d - 2h)b_{trail} + (b_{lead}d + v_{lead})^2} \tag{3.16}$$

3.3.5 Handling oversaturation

Chapter 4

Experimental setup

4.1 Isolated fixed time

Although it may be obvious to expect better solutions by considering individual movements it is interesting to see the degree of superiority.

4.2 Parameters

Cars, trucks and pedestrians.

4.3 Saturation flow rate estimation

Given any pair of semi-compatible movements we can estimate the reduced saturation flow rate by assuming green indication for both movements.

Bibliography

[1] R. B. Allsop. "SIGCAP: A computer program for assessing the traffic capacity of signal-controlled road junctions". In: *Traffic Eng. Control* 17 (1976), pp. 338–341.

[2] R. B. Allsop. "SIGSET: A computer program for calculating traffic capacity of signal-controlled road junctions". In: *Traffic Eng. Control* 12 (1971), pp. 58–60.

[3] S. Alireza Fayazi and Ardalan Vahidi. "Crowdsourcing Phase and Timing of Pre-Timed Traffic Signals in the Presence of Queues: Algorithms and Back-End System Architecture". In: 17 (Nov. 2015), pp. 870–881.

[4] G. Improta and G.E. Cantarella. "Control system design for an individual signalized junction". In: *Transportation Research Part B: Methodological* 18.2 (1984), pp. 147–167. ISSN: 0191-2615. DOI: https://doi.org/10.1016/0191-2615(84)90028-6.

[5] Y. Lin et al. "Transit signal priority control at signalized intersections: a comprehensive review". In: *Transportation Letters* 7.3 (2015), pp. 168–180. DOI: 10.1179/1942787514Y.0000000044. eprint: http://dx.doi.org/10.1179/1942787514Y.0000000044.

[6] John D. C. Little. "The Synchronization of Traffic Signals by Mixed-Integer Linear Programming". In: *Operations Research* 14.4 (1966), pp. 568–594. ISSN: 0030364X, 15265463. URL: http://www.jstor.org/stable/168720.

[7] NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM (NCHRP). *Signal Timing Manual*. 500 Fifth Street, NW Washington, DC 20001: Transportation Research Board Business Office, 2015.

[8] M. Papageorgiou et al. "Review of road traffic control strategies". In: *Proceedings of the IEEE* 91.12 (Dec. 2003), pp. 2043–2067. ISSN: 0018-9219. DOI: 10.1109/JPROC.2003.819610.

[9] S. Samra, A. El-Mahdy, and Y. Wada. "A Linear Time and Space Algorithm for Optimal Traffic-Signal Duration at an Intersection". In: *IEEE Transactions on Intelligent Transportation Systems* 16.1 (Feb. 2015), pp. 387–395. ISSN: 1524-9050. DOI: 10.1109/TITS.2014.2336657.

[10] Martin Treiber. *Traffic Flow Dynamics*. Jan. 2013. ISBN: 978-3-642-32459-8.

[11] F. V. Webster. "Traffic Signal Settings". In: 39 (1958).

Index