

# INTERSECTION TRAFFIC ENGINEERING

Dr John D Sampson

Pr Eng, Fellow SAAE, Fellow SAICE

C 083 307 4666, W 011 231 2298

sampsonjds@gmail.com

sampsonj@ttaafrica.co.za

**Key Words:** intersection control devices: priority control, roundabouts, traffic signals; capacity, delay, queues, level of service; traffic counts; traffic signal staging, phasing and timing.

## SYNOPSIS

This report deals with the practical design of intersections and considers every aspect of traffic engineering that affects Intersection Control Devices including Daily Traffic Volume Patterns, Safety, Capacity, Delay, Queues and Levels of Service.

In developing an intersection optimization and traffic signal design software program called AutoJ, the author did extensive studies of traffic patterns as well as computer simulations of various formulae of capacity and delay for different control devices. The theory which best replicated actual driving behaviour is discussed, but it was found necessary to make modifications to established theory in some instances and to derive new theory in others. The results of these findings and modifications made are reproduced in this report.

These findings are then applied to the design of intersections to determine what traffic control device should be installed. The process required to design a traffic signal cycle, staging and splits for optimal operation throughout the day is also detailed.

# CONTENTS

<b>INTERSECTION TRAFFIC ENGINEERING .....</b>	<b>1</b>
<b>SYNOPSIS .....</b>	<b>1</b>
<b>DEFINITIONS .....</b>	<b>6</b>
<b>ABBREVIATIONS .....</b>	<b>11</b>
<b>1. INTRODUCTION .....</b>	<b>12</b>
<b>2. BASICS.....</b>	<b>13</b>
<b>2.1. Why Intersections are Critical .....</b>	<b>13</b>
2.1.1. Capacity .....	13
2.1.2. Delay.....	13
2.1.3. Crashes.....	14
2.1.4. Summary.....	15
<b>2.2. Intersection Control Devices (ICD) .....</b>	<b>15</b>
2.2.1. Priority Controls .....	15
2.2.2. Traffic Signals .....	16
2.2.3. Traffic signal terms.....	16
2.2.4. Traffic signal stages.....	16
2.2.5. Leading and lagging right turn stages.....	18
2.2.6. The Right Turn, or Yellow, Trap.....	18
2.2.7. The advantages of lagging versus leading right turn flashes.....	19
2.2.8. Permitted and protected phases .....	20
<b>2.3. Choosing an Intersection Control Device (ICD).....</b>	<b>20</b>
2.3.1. Overview of the merits of ICD's .....	21
2.3.2. Warrants for ICD's .....	22
<b>2.4. Putting it all together .....</b>	<b>24</b>
2.4.1. Reasons for an evaluation of the ICD.....	24
2.4.2. Process of evaluation .....	24
<b>3. TRAFFIC VOLUMES .....</b>	<b>26</b>
<b>3.1. Traffic Count .....</b>	<b>26</b>
<b>3.2. Daily traffic patterns .....</b>	<b>26</b>

3.3.	Event Table for traffic signal control .....	29
3.4.	Average Daily Traffic (ADT).....	30
3.5.	Weekly and monthly traffic variations.....	30
3.6.	Lane Volume Variations .....	31
3.7.	Equivalent Vehicle Units .....	32
4.	CAPACITY OF PRIORITY CONTROL DEVICES.....	33
4.1.	Introduction .....	33
4.2.	Two-way stop capacity .....	33
4.2.1.	Derivation of $V_c$ the conflicting flow.....	33
4.2.2.	Derivation of $t_c$ the critical gap .....	34
4.2.3.	Derivation of $t_f$ the follow up time .....	37
4.2.4.	Multiple Approach Lanes .....	38
4.2.5.	Stop street capacity calculation summary .....	38
4.3.	All-Way Stop Capacity .....	39
4.4.	Roundabout Capacity .....	41
4.5.	Mini-Circle Capacity.....	44
4.6.	Pedestrian Capacity .....	44
4.7.	Results .....	44
4.8.	Conclusion on Capacity of Priority Controls .....	46
5.	CAPACITY OF TRAFFIC SIGNALS .....	47
5.1.	Introduction .....	47
5.2.	Capacity Considerations.....	47
5.2.1.	Saturation flow rate .....	47
5.2.2.	Turn caution.....	48
5.2.3.	Grade .....	48
5.2.4.	Pedestrian interference .....	48
5.2.5.	Pedestrian capacity .....	49
5.2.6.	Right turns in gaps at signals .....	49
5.2.7.	Right turns during inter-green .....	49

5.2.8. Effective Green Time .....	50
<b>6. DELAY AND QUEUES .....</b>	<b>51</b>
6.1. Introduction .....	51
6.2. Symbols .....	51
6.3. Delay at Priority Junctions .....	52
6.4. Delay at Under-Saturated Traffic Signals .....	53
6.4.1. Uniform delay .....	53
6.4.2. Random delay .....	54
6.4.3. Correction factor .....	54
6.4.4. Final formula .....	55
6.5. Delay at Oversaturated Intersections .....	55
6.5.1. Modified Bureau of Public Works (BPR) .....	55
6.5.2. Modified Davidson .....	55
6.5.3. JHK .....	55
6.5.4. AutoJ (Sampson 2016) .....	56
6.6. Result comparison .....	57
6.7. Conclusion of Delay Equations .....	62
6.8. Queues .....	62
<b>7. LEVEL OF SERVICE .....</b>	<b>63</b>
7.1. Level of Service range .....	63
7.2. Measures of Effectiveness .....	63
7.3. Performance Index .....	65
7.4. Conclusion .....	66
<b>8. DESIGN OF TRAFFIC SIGNAL TIMINGS .....</b>	<b>67</b>
8.1. Introduction .....	67
8.2. Cycle times .....	67
8.3. Minimum green, yellow and all-red times .....	67
8.3.1. Minimum greens .....	67
8.3.2. Acceleration and deceleration rates .....	68

8.3.3. Minimum yellow and all-red .....	68
<b>8.4. Number of stages .....</b>	<b>70</b>
<b>8.5. Signal timing design .....</b>	<b>70</b>
8.5.1. Overflow vehicles.....	70
8.5.2. Effective number of lanes and lane balance .....	71
8.5.3. Mixed Lanes .....	72
8.5.4. Opposing volume adjustment .....	72
<b>8.6. Critical lanes .....</b>	<b>73</b>
8.6.1. Volume per phase .....	73
8.6.2. Volume per lane.....	74
8.6.3. Optimization method .....	74
<b>8.7. Weighting Favoured Movements .....</b>	<b>75</b>
<b>8.8. Pedestrian Signals .....</b>	<b>75</b>
Requirement for Pedestrian Heads .....	75
8.8.1. Restricted Green for Pedestrians .....	76
8.8.2. Early Start for Pedestrians .....	76
<b>8.9. Co-Ordinated Fixed Time or Vehicle Actuation .....</b>	<b>77</b>
<b>MAIN REFERENCES.....</b>	<b>78</b>

# INTERSECTION TRAFFIC ENGINEERING

## DEFINITIONS

The definitions of terms that are used throughout, some of which are specifically formulated for this report, are listed here.

ar	<p><b>All-red</b> - the <u>phase</u> (portion of a <u>stage</u>) during which all the approaches to the intersection are shown a red signal. It is the safety clearance period which allows traffic to clear the intersection before opposing vehicles are given a green signal.</p> <p>In some countries red and yellow signals are displayed simultaneously during the “all-red” period as an early warning to the approaches about to be given a green signal, but this is not permitted in South Africa.</p>
Access	<p><b>Access</b> - any public or private road, ramp, driveway, intersection or path which crosses or is connected to a public road.</p>
AADT	<p><b>Average Annual Daily Traffic</b> – the total traffic volume in a year, including school and public holidays and weekends, divided by 365.</p>
ADT	<p><b>Average Daily Traffic</b> – the 24-hour traffic count taken on a typical week day in an urban area (see <u>Daily Variation</u>). As weekends and holidays are excluded, ADT is greater than <u>AADT</u>. ADT should be used in preference to <u>AADT</u> when doing urban traffic studies.</p>
AWDT	<p><b>Average Week Day Traffic</b> – the total traffic in a week without school or public holidays divided by five. AWDT is approximately equal to <u>ADT</u> in an urban area.</p>
AM (PM)	<p><b>Ante Meridiem (Post Meridiem)</b> - the <b>morning</b> (a.m.) or <b>evening</b> (p.m.) peak hour or peak period, depending on the context. Traffic signals are normally set for a three-hour peak period based on the peak one-hour traffic volume.</p>
AWS	<p><b>All-way Stop</b> control – commonly known as the 4-way Stop but could be 3-way at a T-junction.</p>
C	<p><b>Capacity</b> (vehicles per hour) - the maximum volume that can be handled by the control device. For signals, nominally <math>C = \underline{S} \times (\underline{g/c})</math>, but more correctly is adjusted for vehicle composition (e.g. heavy vehicles and buses), geometric considerations (e.g. grade, lane numbers and widths, turning penalties, pedestrian interference), start-up lost time and inter-green overruns.</p>
μ	<p><b>Co-ordination</b> factor - a factor was developed specifically for use in AutoJ and varies from 0% to 100%. It can be directly related to the Highway Capacity Manual</p>

arrival types (TRB 2000). The co-ordination factor can be considered as the percentage of all vehicles that arrive during the green phase:

- **100%** when **all** vehicles, whether in a platoon or evenly spaced, arrive during green phase;
- **50%** (default) when there are random or uniform arrivals and vehicles are approximately evenly spaced throughout the cycle;
- **0%** when entire platoon arrives during the red phase (no vehicles arrive during the green phase).

CLV	<b>Critical Lane Volume</b> (veh/hr/lane) - the equivalent volume in the lane which has the greatest utilization of <u>capacity</u> , i.e. the lane “busiest” that needs the longest green time.
c	<b>Cycle</b> time (seconds) - the time to go from the start (or end) of any <u>phase</u> , back to the start (or end) of that same phase again. Fixed time signals have fixed cycle times; actuated signals vary the time of each cycle.
DV	<b>Daily Variation</b> - the variation in traffic volume from day to day. Monday to Sunday traffic volumes as a percentage of AWDT are respectively 97, 98, 100, 100, 105, 73, 53 (Sampson 1983, 2017) but also vary week by week. Despite the variation, it is generally acceptable to do a count on any typical weekday.
d	<b>Delay (ave)</b> – average additional travel time, in seconds, resulting when vehicles must slow down and/or stop for any reason. Delay is incurred decelerating, stopping, waiting to be served (waiting to cross the stop line, including delay while moving up in the queue) and accelerating. Most research studies exclude deceleration and acceleration delays for signals but include them for Stop and Yield controls.
D	<b>Delay (tot)</b> – total delay, in vehicle-hours per hour, experienced by vehicles that are delayed in any way, even if not stopping. If acceleration and deceleration delay is excluded, total delay exactly equals queue length numerically. In AutoJ, total delay and queue are taken to be equal even though this slightly overestimates queues at priority controls.
drive	<b>Driveway</b> - a roadway giving access to property.
% grade	<b>Grade</b> – down (minus) or up (plus) slope on the approach to the stop line, expressed as a percentage; -2% to +2% is flat, 2% to 5% is a gentle incline, 5% to 8% is a steep incline and greater than 8% is very steep, seldom found.
g	<b>Green</b> - duration of green <u>phase</u> (in seconds) of any <u>stage</u> .

g/c	<b>Green to Cycle</b> time ratio - percentage of green available to an approach.
y + ar	<b>Inter-green</b> - interval or time between the end of one green <u>phase</u> and the start of the next one, consisting of the yellow plus all-red <u>phases</u> .
i/s	<b>Intersection</b> - the point at which two public roads join or cross at grade (see <u>Access</u> ).
ICD	<b>Intersection Control Device</b> - the device or method by which the intersection is controlled. The ICD can be a priority (stop, yield and roundabout) or a traffic signal control. Uncontrolled <u>intersections</u> and <u>driveways</u> have no ICD and operate based on gap acceptance.
LOS	<b>Level of Service</b> - ease of movement rated from A (free flow) to F (excessively congested, demand exceeds capacity) in accordance with the Highway Capacity Manual (TRB 2010).
off	<b>Off-peak</b> – the period between the AM and PM peak periods, normally with approximately equal flows in both directions. In urban areas, nights and weekends are also “off-peak” and typically use the same signal settings as for weekday off-peak periods. “Unusual” peaks, such as at shopping centres on Saturdays or sports stadia before and after matches, might not be typically off-peak and the signals may need to be reset accordingly.
pk hr	<b>Peak hour</b> – the highest hour/s during the day. In urban areas there are two peak hours, AM and PM. The urban AM peak hour generally occurs between 6:45 and 7:45 and the PM peak between 16:30 and 17:30. These are the busiest hours city-wide, even if they are not the hours with the highest traffic flow at a specific location.
pk	<b>Peak periods</b> – the duration of high traffic volumes in urban areas, usually longer than the peak hour. Traffic signals generally start the peak programme 30 minutes before the peak period (to prevent build up) and end 30 minutes after (to allow overflow), hence 6:00 to 9:00 and 15:30 to 18:30 are typically used for peak period programmes.
ph	<b>Phase</b> – the signal display at any point in a cycle. A sequence of green, yellow and all-red phases combined constitute a <u>stage</u> . Strictly speaking the “all-red” period is not a phase on its own but an overlap of the red phases, but it is not uncommon to refer to an all-red phase. Pedestrian crossing phases (green man, flashing red man and steady red man), if provided, can be displayed at the same time as vehicle phases, or they can be displayed as a separate <u>stage</u> (scramble stage).



Q	<b>Queue</b> length - the total number of vehicles stopped in a lane or waiting to enter an intersection at any point in time. Vehicles approaching the back of the queue and vehicles leaving from the stop line at the front of the queue are not in the queue, but vehicles moving up within the queue are in the queue.
QED	<b>Quod E'rat Demonstrandum</b> - as proven, loosely translated as Quite Easily Done.
r	<b>Red</b> - the period on an approach where the signal is not showing green or yellow.
RR/CC	<b>Roundabout</b> - an intersection, arguably including a mini-circle, where all approaching vehicles must yield to vehicles in the circulating roadway. Vehicles approaching a roundabout or mini-circle give way to conflicting vehicles on their right who are already in the circular roadway (even if those vehicles entered the circulatory roadway after a vehicle waiting at the yield line) but not to those that have not yet crossed the yield line. If any approach is controlled in any way other than having to yield, or if the circulatory roadway is not free running, the intersection is not a roundabout, even if it is circular in shape.
RTSM	South African / Southern African Development Communities <b>Road Traffic Signs Manual</b> , as in SARTSM (Volumes 2 and 3) or SADC RTSM (Volumes 1 and 4).
S	<b>Saturation</b> flow - the maximum flow rate in vehicles per hour at which an “infinite” queue of light vehicles can flow passed a point in a single lane of traffic in normal circumstances (flat road, good visibility, and adequate width paved lane). A saturation flow rate of 2 000 veh/hr is taken for AutoJ.
stg	<b>Stage</b> – a specific combination of green, yellow and “all-red” <u>phases</u> displayed to an approach or approaches simultaneously. Often called a phase in South Africa, as in “3 phase” or “4 phase” on traffic signals boards, but stage is the preferred term internationally. The minimum and simplest traffic signal setting is two stages (2), made up of six <u>phases</u> .
T	<b>Total approach (through)</b> traffic, consisting of left (L), straight (S) and right (R) turn traffic proceeding on the same green disc.
traf	<b>Traffic</b> – vehicles, motorcycles, bicycles, pedestrians and any other form of conveyance using a road. Commonly, traffic is used to refer to vehicles only, e.g. average daily traffic.
TCD	<b>Traffic Control Device</b> - a sign, road marking or traffic signal (robot) controlling the behaviour of vehicles. (The Manual of Uniform <i>Traffic Control Devices</i> (MUTCD) is the USA equivalent to the SA RTSM.)

TWSC	<b>Two-Way Stop Control</b> - a form of control where one side is stopped and the other side given priority (not stopped), hence also is known as priority control.
V	<b>Volume</b> (veh/hr) - the number of vehicles passing a point in a fixed period, typically one hour. Cars, trucks, buses and taxis are “vehicles” while motorcycles, bicycles and pedestrians are normally not counted in this total, and if counted, it is necessary to specify such. Ideally “volume” should be the demand volume, but usually is the counted volume. The demand volume is equal to the counted volume plus any queue waiting at the beginning and end of the counting period.
V/C	<b>Volume to Capacity ratio</b> , equal to the <u>Volume</u> divided by <u>Capacity</u> (Degree of Saturation is the <u>Volume</u> to <u>Saturation</u> flow ratio).
q	<b>Volume</b> (veh/sec) or traffic flow in <i>vehicles per second</i> (see V, veh/hr).
y	<b>Yellow</b> , or amber, phase of any stage.

## ABBREVIATIONS

Code	Type
0	No control
Xns, Xwe, Yns, Ywe	X = Stop, Y = Yield; ns = control on north-south approaches (west-east run free), and vice versa.
XX	All-way stop (4-way stop)
mC, RR	Mini-circle, Roundabout
2	2 stage signal (north-south through together then west-east through together)
3ns, 3we	3 stage signal, 2 plus lagging green from north and south, or west and east
n3, s3, w3, e3	3 stage signal, 2 plus leading green from north, south, west, east
4nswe	4 stage signal, 2 plus lagging green from north and south and lagging green from west and east
n4we, s4we, w4ns, e4ns	4 stage signal, 2 plus leading green from north plus lagging green from west and east, etc.
nw4, ne4, sw4, se4	4 stage signal, 2 plus leading green from north plus leading green from west, etc.
n-s-3, w-e-3	3 stage signal, split north only, then south only, then west-east together, etc.
n-s-4we, w-e-4ns	4 stage signal, split north only, then south only, then west-east, then lagging w-e, etc.
n-s-w-e-4	4 stage signal, split north only, south only, west only, then east only

# 1. INTRODUCTION

This report contains the findings and application by the author of international best practice for the design and optimization of individual intersection control devices.

It serves three purposes:

1. A practical intersection design manual;
2. Course notes for an intersection design course;
3. A technical manual for the software intersection design program AutoJ.

Subjects covered in the report include the following Chapters:

1. This Introduction
2. Basics (Practical Intersection Design Tips)
3. Traffic Volumes
4. Capacity of Priority Controlled Intersections
5. Capacity of Traffic Signalized Intersections
6. Delay and Queues
7. Level of Service
8. Design of Traffic Signals

And two Appendices

- A. Leading and Lagging Flashing Green Arrows
- B. The Challenge of All-Way Stops
- C. References Current
- D. References Older

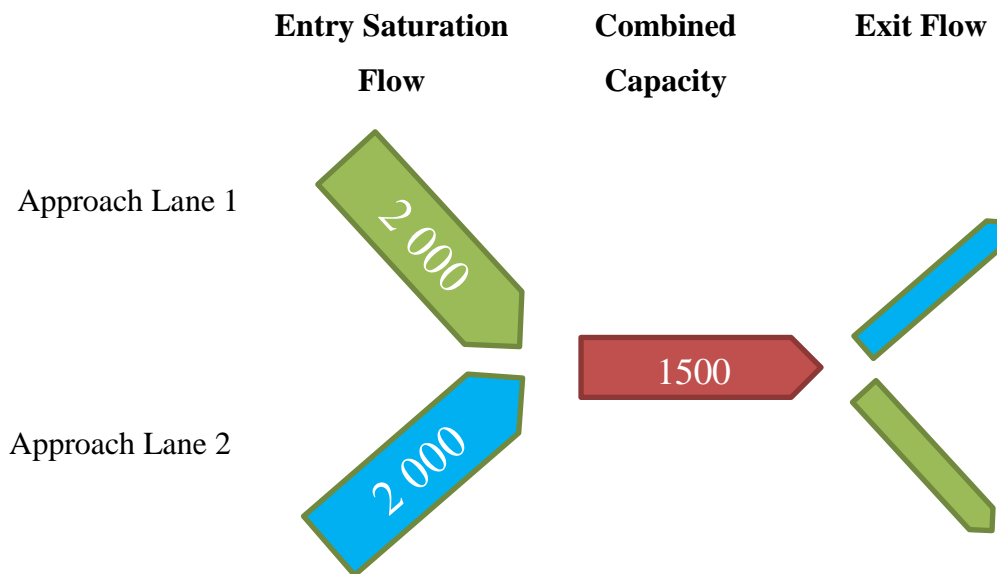
## 2. BASICS

### 2.1. WHY INTERSECTIONS ARE CRITICAL

The intersection is where most of the problems associated with traffic flow will occur. It is the place with the least **capacity**, the most **delay** and the highest **crash** rate. Each of these will be dealt with in detail later, but the concepts are explained below.

#### 2.1.1. CAPACITY

At intersections, **Capacity** is squeezed, as illustrated below (veh/hr):



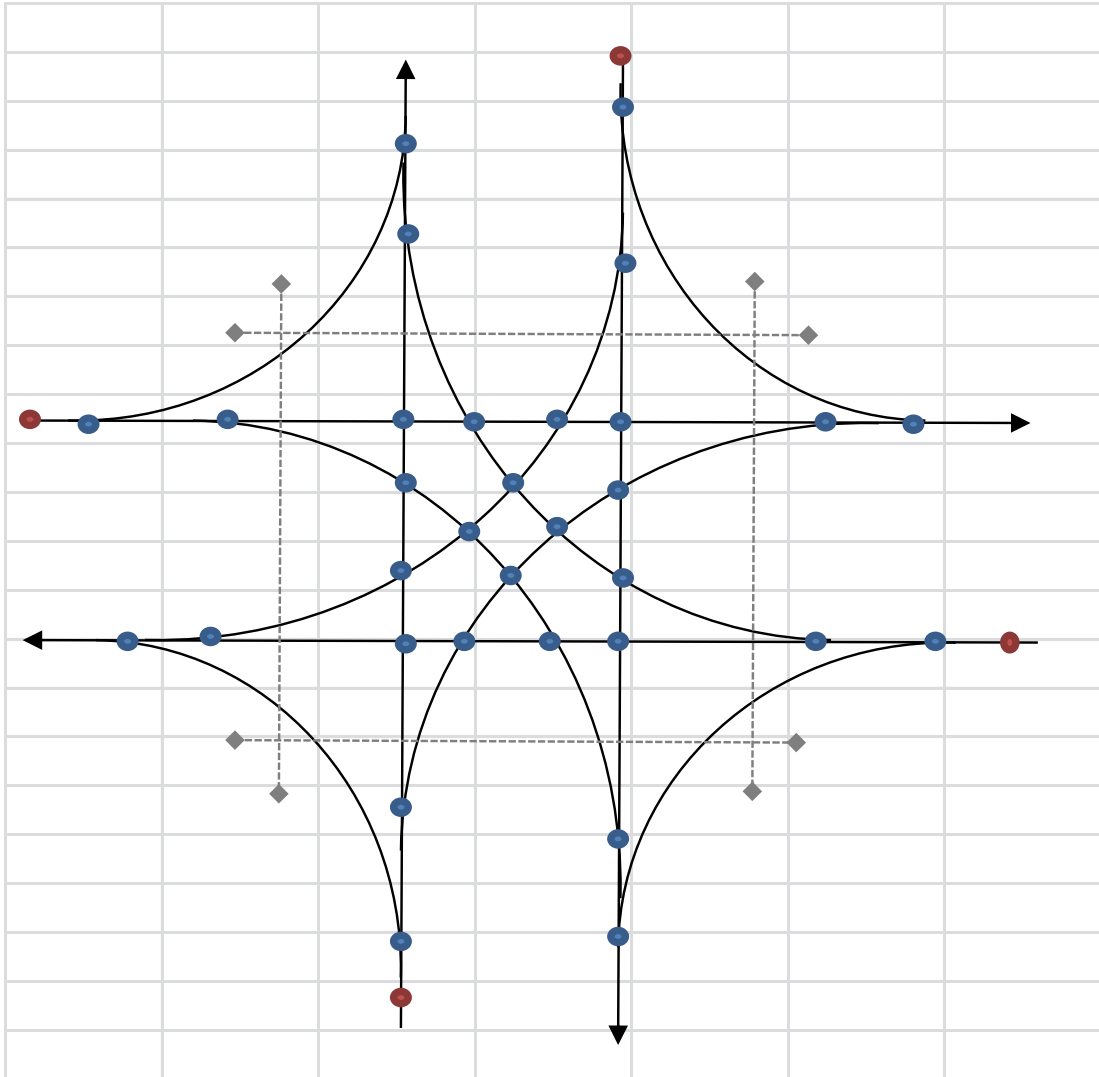
The saturation flow of an unimpeded lane is approximately 2 000 veh/hr. Two approach lanes could therefore carry up to 4 000 veh/hr before joining. After they meet and interact, it is only possible for around 1 500 vehicles per hour or less to exit. This 1 500 must now be shared between the two lanes based on the demand volumes.

#### 2.1.2. DELAY

At intersections, **Delay, stops and queues** are always incurred. Even under light flow conditions at least one approach to the intersection will have to give way at times.

### 2.1.3. CRASHES

**Crashes, or accidents**, are at their highest at intersections. Even without cyclist and Bus Rapid Transit lanes being added, at every cross-junction there are 32 **vehicle** conflict points (blue) and 24 **pedestrian** conflict points, illustrated in Figure 1 below. In addition, there will be the stop-start conflict (red) caused by the introduction of an ICD.

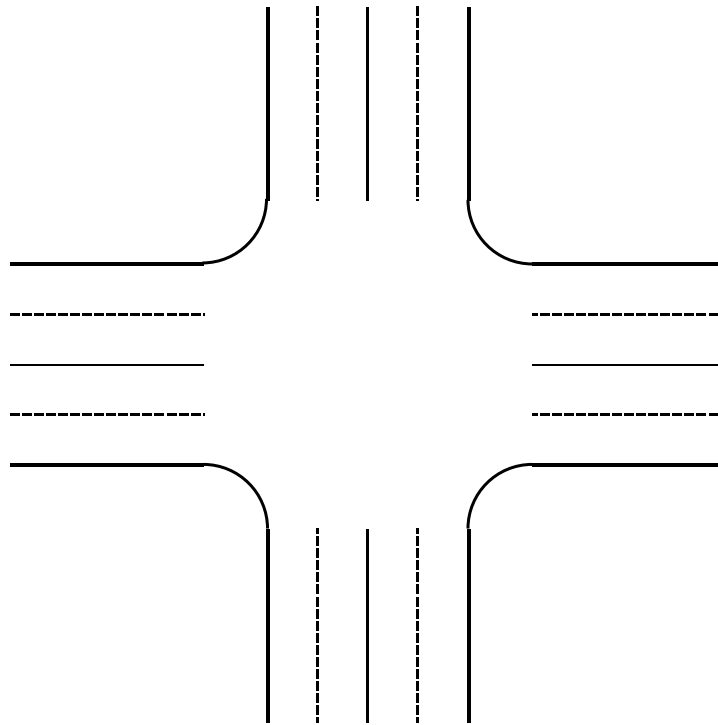


**Figure 1: Vehicle conflict points with pedestrian crossings shown dashed**

Furthermore, the complexity and decision making at intersections greatly increases the **load and stress** on both drivers and NMT (non-motorized pedestrians and cyclists).

#### 2.1.4. SUMMARY

A typical intersection may look as follows:



The traffic engineer's job will be to decide the road markings and control device that will maximize the capacity, minimize the delays, maximize the safety and optimize the Level of Service under all operating conditions.

### 2.2. INTERSECTION CONTROL DEVICES (ICD)

There are two basic types of intersection control, 1) **priority** control or 2) traffic **signal** control.

#### 2.2.1. PRIORITY CONTROLS

**Priority controls** (the six bullets below) are all fixed once implemented and cannot be adjusted for time of day or varying traffic conditions.

- No control
- Yield (on one side)
- Mini-circle (Yield on all sides)
- Roundabout (Yield on all sides)
- Stop (on one side)
- All-way or 4-way Stop (Stop on all sides)

### 2.2.2. TRAFFIC SIGNALS

**Traffic Signals** on the other hand are highly flexible and have literally an infinite number of possible operational settings. Traffic signals can be varied by time of day, day of week or in the case of **vehicle actuated** signals, continuously. They can have different **cycle** times, different **staging**, different **splits** and can be co-ordinated, or synchronized, with other traffic signals.

### 2.2.3. TRAFFIC SIGNAL TERMS

The sketch below illustrates the concepts of cycle, stages and phases.



In this example there are two **stages**, 1 and 2. Each stage is made up of three **phases**, a green phase, a yellow phase and a red phase. The **all-red** “phase” is the time during which the two red phases overlap.

The **cycle time** is equal to stage 1 plus stage 2. The **split** is the time given to each stage, in this case 50:50. The combined yellow and all-red time is known as the **inter-green**.

These concepts are also illustrated on page 6.2 of the SA RTSM Manual Volume 3. Note however that in the Manual the concepts are slightly confused in that the illustration of a stage does not include the inter-green and therefore appears to be exactly like a phase.

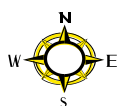
### 2.2.4. TRAFFIC SIGNAL STAGES

There are three legally permitted types of signal staging illustrated on the following page:

1. A **main** or through stage, where all movements on the two opposite approaches are given green at the same time;
2. A **single right turn flash** stage where the left and straight disc, and right turn arrow movements from the same approach are displayed at the same time;
3. A **double right turn flash** stage, where opposite right turns flash at the same time.

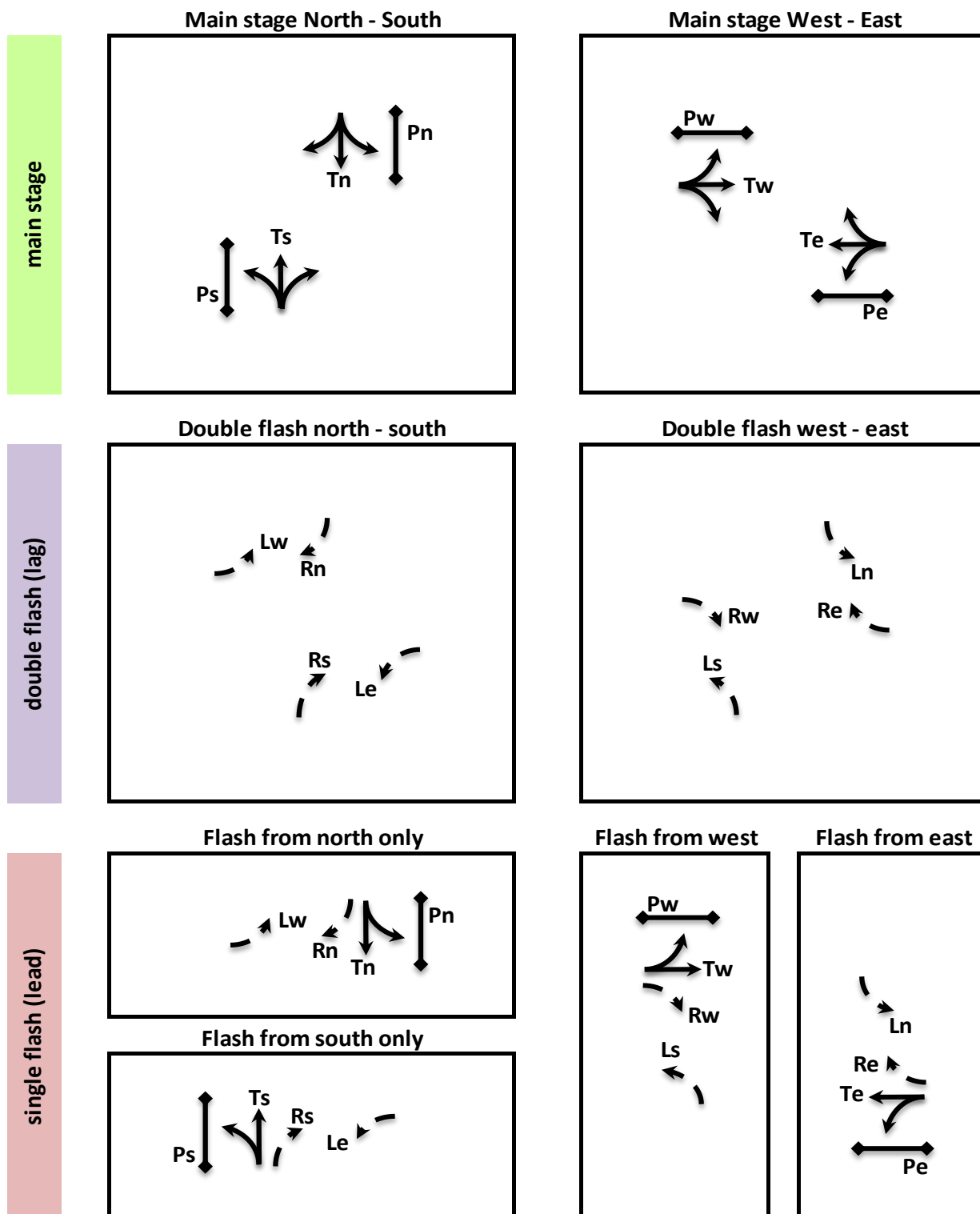
This gives us eight potential ways of displaying signal stages which can be displayed in any order. Theoretically all eight could be used at a single intersection. The scramble stage, where all vehicular traffic is stopped, and pedestrians can proceed in any direction, is a potential additional option.





## Legally Permitted Signal Combinations

*Movement(s) and stage(s) may be omitted, but none may be added.  
Stages may be displayed in any order. Minimum two, maximum eight, prefer fewer.*



**Key** T = Through, steady disc display  
 R = Right turn, flashing arrow display  
 L = Left turn, flashing arrow display  
 P = Pedestrian, green man + flashing red man display

n s w e = North, South, West, East  
 solid line = permitted  
 dashed line = protected

**Figure 2: Legally permitted stages**

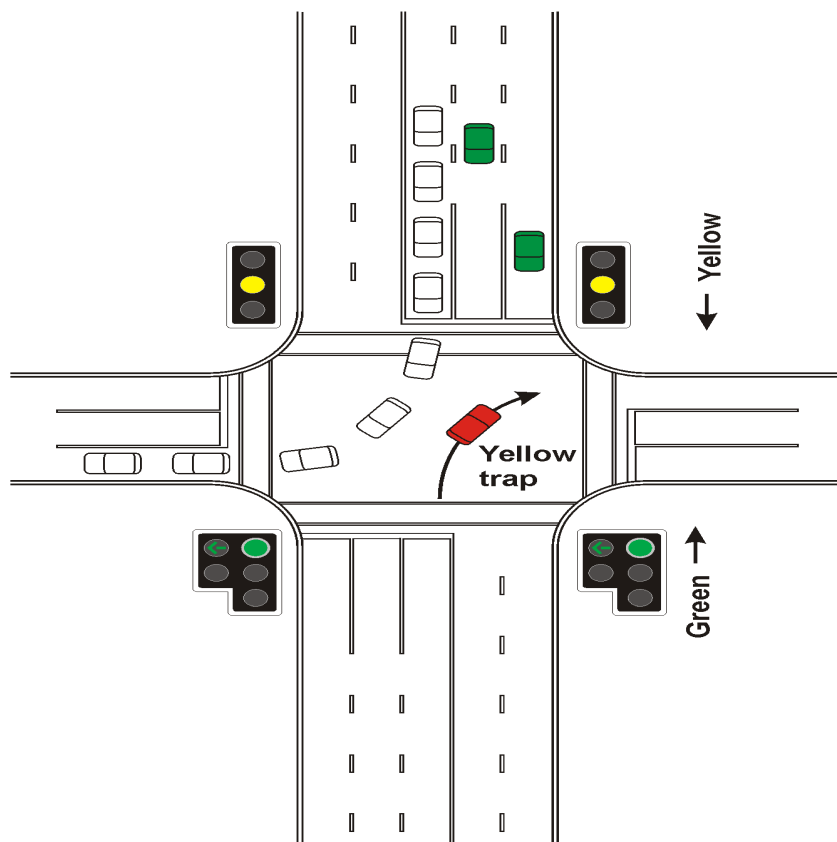
### 2.2.5. LEADING AND LAGGING RIGHT TURN STAGES

If a flash is displayed before the **main** stage, it is described as **leading**, if after it is **lagging**. If the signal gives green to only one approach at a time (single flash in Figure 2, not followed by main stage), it is known as a **split** stage.

Given that stages can be displayed in any order, this section discusses whether some sequences are better than others. For reasons given in Appendix A and in 2.2.6 and 2.2.7 below, the recommended rule for choosing between lagging and leading green arrows is as follows:

- **Leading:** If the flash is needed from one side only and the opposite right turn is possible, the flash **must** always be leading (lagging is not allowed in this situation). The through green should always appear at the same time as the leading arrow on that approach.
- **Lagging:** In all other cases lagging arrows are more efficient and safer; and therefore preferred. Lagging right turn flashes therefore apply for double right turns, single right turns at T-junctions or right turn flashes on one-ways where the opposite right turn is not possible.

### 2.2.6. THE RIGHT TURN, OR YELLOW, TRAP



**Figure 3: The Right Turn, or Yellow, Trap**

The **right turn trap** occurs when turning is allowed from both sides of the intersection as in Figure 3, and a single lagging green flash is displayed.

In Figure 3, the right turn volume from north (nR) greatly exceeds the right turn from south (sR), so the flash is needed on the north side only. Furthermore, the through traffic from north (nT) is busier than from south, so nT runs at the same time as nR.

However, consider the right turner from south (red in the Figure). At the end of the main green phase, this vehicle is faced with a yellow then red light but may not go because the light is still green for the opposite side. The vehicle is “trapped” in the intersection and there is the real danger that the vehicle will turn in the face of oncoming traffic.

This signal sequence (**single lagging green flash** when opposing right turn is possible) is therefore not permitted.

Hence the first bullet point rule is 2.2.5.

#### 2.2.7. THE ADVANTAGES OF LAGGING VERSUS LEADING RIGHT TURN FLASHES

If the right turn trap is not a concern, then generally the flash should be lagging (second bullet point rule in 2.2.5) for the following reasons:

- **Lagging** green right turn arrows:
  - Comply with the Rule of the Road
  - Meet user expectations
  - Eliminate hazardous late turns in the face of oncoming traffic
  - Reduce pedestrian and vehicle false starts
  - Separate right turning and pedestrian conflicts
  - Improve road safety
  - Improve capacity
  - Improve efficiency of vehicle actuated right turns.

However, if the flash is only needed from one side, then leading is better for the following reasons:

- **Leading** green right turn arrows:
  - Avoid the right turn trap
  - Cater for unbalanced or tidal flows
  - Cater better for shared lanes and short auxiliary lanes
  - Reduce gap acceptance conflicts
  - Are better if stage is to be skipped when not needed.

For further motivation, Annexure A discusses leading and lagging flashes in detail.

#### 2.2.8. PERMITTED AND PROTECTED PHASES

Where a separate right turn flash stage is not installed, a driver turning right can take gaps or turn during the inter-green period. The right turn in gaps and during inter-green is therefore **permitted**.

When a separate right turn flash stage is provided, and drivers utilize this stage to make the right turn, they have a **protected** right turn as all conflicting movements are prohibited.

In cases where the driver has the choice of taking a gap or using the right turn flash, this is the **permitted/protected** signal phasing.

Where turning in gaps is prohibited by a red signal and ST sign, drivers may only turn during the right turn flash stage. This is the **protected-only** signal phasing.

There are no absolute rules for when permitted or protected staging should be applied. “**Protected only**” is however poorly understood and poorly obeyed by the travelling public and the efficiency and capacity of the intersection is reduced by this restriction.

Therefore, it is sensible to only use “**protected only**” when right turning traffic cannot see vehicles approaching from the opposite direction due to geometric sight line obstructions (crest vertical curves or bends in the road), or where three or more lanes turn simultaneously. They are also necessary when there are BRT lanes down the median.

The Gauteng Department of Public Transport, Roads and Works report BL03/01: “**Double right turn lanes on Provincial Arterials**”, April 2003 provides the following guidelines:

**Table 1: Permitted and protected phasing guidelines**

Right-turn volume per hour	Number of right-turn lanes	Right-turn signal phasing
0 – 120	1	Permitted
120 – 300	1	Permitted/Protected ( <i>or Protected Only</i> )
300 – 600	2	Permitted/Protected ( <i>or Protected Only</i> )
> 600	3	Protected Only

### 2.3. CHOOSING AN INTERSECTION CONTROL DEVICE (ICD)

Given the various types of ICD, how do we decide which one is best? To determine this, we must first decide what we mean by best.

Normally we will want to achieve the following:

- Maximum capacity;
- Minimum delay;
- Shortest queues;
- Maximum safety;
- Minimum cost and maintenance.

Unfortunately, these goals are often in conflict. We therefore need to decide which we consider most important which in turn will depend on the traffic conditions and the strengths and weaknesses of each type of control.

### 2.3.1. OVERVIEW OF THE MERITS OF ICD'S

Under ideal conditions, the merits of each ICD, numbered from best to worst, is as follows:

#### **Capacity**

1. Traffic Signal
2. Roundabout
3. Stop / Yield
4. 4-way Stop / mini-circle

#### **Delay and Queues**

1. Mini-circle (low volume)
2. Roundabout
3. Yield
4. Stop
5. Traffic signal
6. 4-way stop / mini-circle (high volume)

#### **Safety**

1. Roundabout
2. Stop / Yield / mini-circle
3. Traffic signal
4. 4-way stop

#### **Cost (life cycle)**

1. Stop / Yield / 4-way stop / mini-circle
2. Roundabout

### 3. Traffic signal

Clearly, roundabouts emerge as the best or near best in most categories, traffic signals are poor in most categories except capacity and 4-way Stops are the all-round worst performers (why all-way Stops are bad for safety is explained in Appendix B).

In the absence of a full analysis therefore, the above gives us a guide on what to consider.

#### 2.3.2. WARRANTS FOR ICD'S

Generalized warrants for each intersection control device are given below. While serving as a guideline, a more detailed analysis should be undertaken before installing traffic signals or a roundabout.

##### 2.3.2.1. NO CONTROL

If a driveway, or lightly trafficked intersection, or within a residential precinct (designated as such by an appropriate sign), **no control** device is necessary. In the absence of a control, pedestrians will have priority, followed by cyclists, and vehicles must give way to non-motorized transport and each other.

##### 2.3.2.2. YIELD (AND STOP/YIELD)

A **yield** sign is an under-utilized control device suitable for a large range of applications. For a yield sign to be safe however there must be adequate sight distance, measured as follows:

- Standing in the minor street approach, measure nine metres back from the stop line (or proposed stop line) and check how far in each direction a vehicle can be seen;
- If sight distance is 180 metres or more in both directions, install a Yield sign;
- If sight distance is 180 metres or more to the right but less to the left, install a Stop/Yield sign.

##### 2.3.2.3. MINI-CIRCLE

A **mini-circle** is suitable in low to middle volume applications, preferably with relatively balanced traffic flows. It is a good traffic calming device and should be considered on any Class 4b residential collector and Class 5b residential local street. In urban residential areas mini-circles should be used to replace all-way stops. They can also be used (with warning signs) where sight distance is poor.

##### 2.3.2.4. STOP

As a rule, two-way stop or yield streets are suitable for low traffic volume situations where the main road is busier than the side street.

If no control, yield control or a mini-circle cannot be justified, install a **Stop** sign. Place the Stop sign on the minor road or on the stem of a T junction. In a 60 km/h area, there must be a minimum sight distance of 100 metres in either direction, clear of vegetation and other obstructions. Provide a splay if necessary.

#### 2.3.2.5. ROUNDABOUTS

Roundabouts become optimal as volumes grow. They operate effectively in both low volume and high-volume situations.

The following is a summary of the guidelines for the location of **roundabouts** extracted from local and international literature and as contained in the BL 99/5 draft report **Roundabouts (Traffic Circles) as Intersection Control Devices on Provincial Roads**, March 2001, PWV Consortium.

The **best** locations for roundabouts are as follows:

- Intersections where **safety** would otherwise be a problem;
- Intersections where **environmental** enhancement or landscaping is required;
- Intersections where traffic signal maintenance or the availability of power is a problem;
- Where signals are required but cannot be afforded;
- Where permanent, **maintenance free** control without enforcement is necessary;
- Where there are four-way or **all-way stops**;
- Where the **road standard or speed limit changes** (e.g. where an arterial road changes to collector/local status);
- Where urban and rural roads meet (e.g. entrance to towns) or where commercial/industrial and residential areas meet;
- In suburban areas where **traffic calming** is required;
- At intersections with **high turning movements**;
- Where three or more stages are required at traffic signals, roundabouts should be considered;
- Where **U-turns** are prevalent or desirable;
- At intersections with more than four legs;
- At Y-junctions or other junctions with awkward geometry (e.g. sharp change in direction);
- With other roundabouts in a network where **intersection spacing is too close** for signal coordination to be achieved.

Roundabouts equalise the priority of all approach roads. No matter how minor the intersecting road may be, it is afforded the same priority on entry as any of the major routes. Furthermore, all vehicles must slow and take gaps on approaching the roundabout and priority cannot be given to

any movement without violating the roundabout operational principles (e.g. once traffic signals or stop streets are installed at roundabouts, they cease to operate as roundabouts).

The **worst** locations for roundabouts are as follows:

- Where main road and side road **traffic flows differ greatly**;
- Where **minor crossroads enter major routes** when a stop or yield street would suffice;
- In **signalised co-ordinated networks** where they would break up the platoon flow;
- Where traffic signals will soon be required.

As is the case of all intersection control devices, roundabouts should be avoided on roads with steep slopes or where the intersection is not visible. Longer ‘flat’ areas are required for roundabouts compared with other intersection types, making them less suitable on steep grades.

#### 2.3.2.6. TRAFFIC SIGNALS

Traffic signals become needed in high volume, multi-lane situations. Traffic signals should only therefore be considered when the priority controls described above do not provide adequate capacity or result in excessive (intolerable) delay.

The 4Q or 6Q warrant for vehicular or pedestrian traffic signals (as described in Volume 3 of the SA Road Traffic Signs Manual, Chapter 2) was developed to identify when the queue length at a Stop street is such that signals are required. Modern roundabouts, which can form a bridge between a Stop street and traffic signal, were not considered in developing the warrant.

#### 2.3.2.7. ALL-WAY (4-WAY) STOP

There is no warrant for an all-way Stop in an urban area. Use a mini-circle or any other ICD instead.

## 2.4. PUTTING IT ALL TOGETHER

### 2.4.1. REASONS FOR AN EVALUATION OF THE ICD

There is normally a reason why an intersection needs to be evaluated. Typical reasons are:

- Complaint received (congestion or safety);
- Request for traffic signal;
- Traffic impact assessment (new development, future projection);
- Change in traffic pattern;
- New construction, including BRT;
- Observation or maintenance program.

### 2.4.2. PROCESS OF EVALUATION



This section contains a summary of the process that the traffic engineer must carry out to properly analyse an intersection. The process is described and then each step is analysed in detail in the Chapters that follow.

In preparation for the analysis, in all cases it is essential to:

- 1) **Visit the site**, get a feel for how it is working, where pedestrians cross, visibility, hawkers etc., things that cannot be seen on Google Earth;
- 2) Note the geometry and **lane markings**;
- 3) Get a **traffic count** (typically during AM, PM and off-peak) of all turning movements and;
- 4) If available, check the **crash record**.

The principle is to then to choose the minimum level of control that can be justified and is safe.

The analysis is then carried out in the following sequence:

1. Analyse the Vehicle **Volume** (and pedestrian plus cyclist volume if available) for a minimum of the AM and PM hours, but preferably off-peak as well. Adjust for heavy vehicles to get equivalent vehicle units (evu);
2. Measure queue lengths, particularly during peak hours;
3. Note the **Geometry**, including lane markings and slip roads, pedestrian crossings, median islands, clearance distances for vehicles and pedestrians, grade, approach speed, auxiliary lane lengths and, if applicable, details of the signal operation;
4. Add additional traffic volumes that may be using the intersection, e.g. due to a development;
5. Determine the control(s) most likely to be needed at the intersection;
6. Use the above values to calculate the **Capacity** of each lane group for each of the ICDs that might be considered;
7. Determine the effective volume using each lane and use that to identify the **Critical Lanes**;
8. If a signal is applicable, calculate the optimal **Signal Timings** for each period;
9. Based on the green times, calculate the **Volume to Capacity ratio (V/C)** for each movement, approach and for the intersection as a whole;
10. Calculate the **Delay** for each movement and determine the maximum and average delay for vehicles and pedestrians;
11. From the volumes and delay, calculate the **Queue** lengths;
12. Based on V/C and delay for each movement, approach and intersection, determine the **Level of Service**;
13. Using a combination of V/C, delay and queue to give a **Performance Index**, determine the best performing ICD during each period.

14. Select the control with the best overall performance;
15. Prepare signal **Timing Diagrams** including green, yellow, all-red, pedestrian green man and flashing red man times for the preferred option(s);
16. **Report** the results.

As can be seen, the procedure to design an intersection requires considerable technical knowledge, skill and time. To help the user, the author has developed a computer software program to automatically carry out all the required steps to a high level of accuracy. That program is called **AutoJ**.

### **3. TRAFFIC VOLUMES**

This Chapter will consider typical traffic patterns and give some practical guidelines on how and when to conduct traffic counts.

#### **3.1. TRAFFIC COUNT**

The first requirement for any control device decision is a traffic count. This count should as a minimum include all straight and turning movements of vehicles using the intersection. Note should also be taken of heavy vehicles, public transport and pedestrians / cyclists. If these latter variables are not able to be estimated, they should also be counted.

#### **3.2. DAILY TRAFFIC PATTERNS**

In most urban areas, the peak traffic periods are generally consistent and foreseeable and are known to motorists. At any time on a normal weekday, the arrival volume and patterns on the approaches to an intersection can be predicted in advance. This is illustrated in Figures 4 and 5.

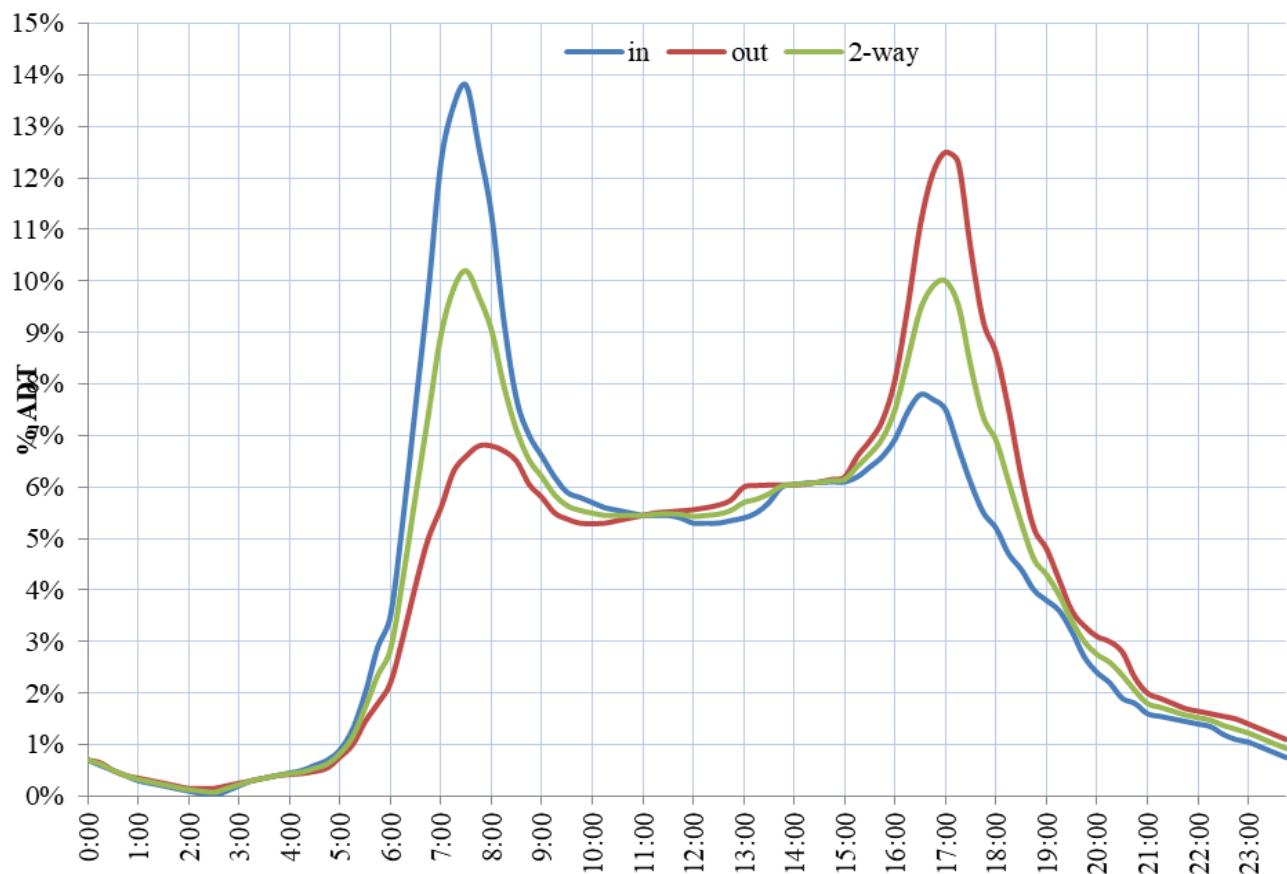
Figure 4 is a compilation of hundreds of 12 and 24-hour counts taken in the 1970's and 1980's, mainly in Johannesburg (Sampson 1983). Figure 5 is derived from eleven arterial and freeway SANRAL Comprehensive Traffic Observations (CTO) locations throughout Gauteng in 2016 (Sampson 2017).

In congested areas and CBDs, the peaks are flatter and last longer while in uncongested areas the peaks are sharper and shorter. In areas remote from employment centres, peaks start earlier in the morning and end later in the afternoon while in CBD's the opposite takes place.

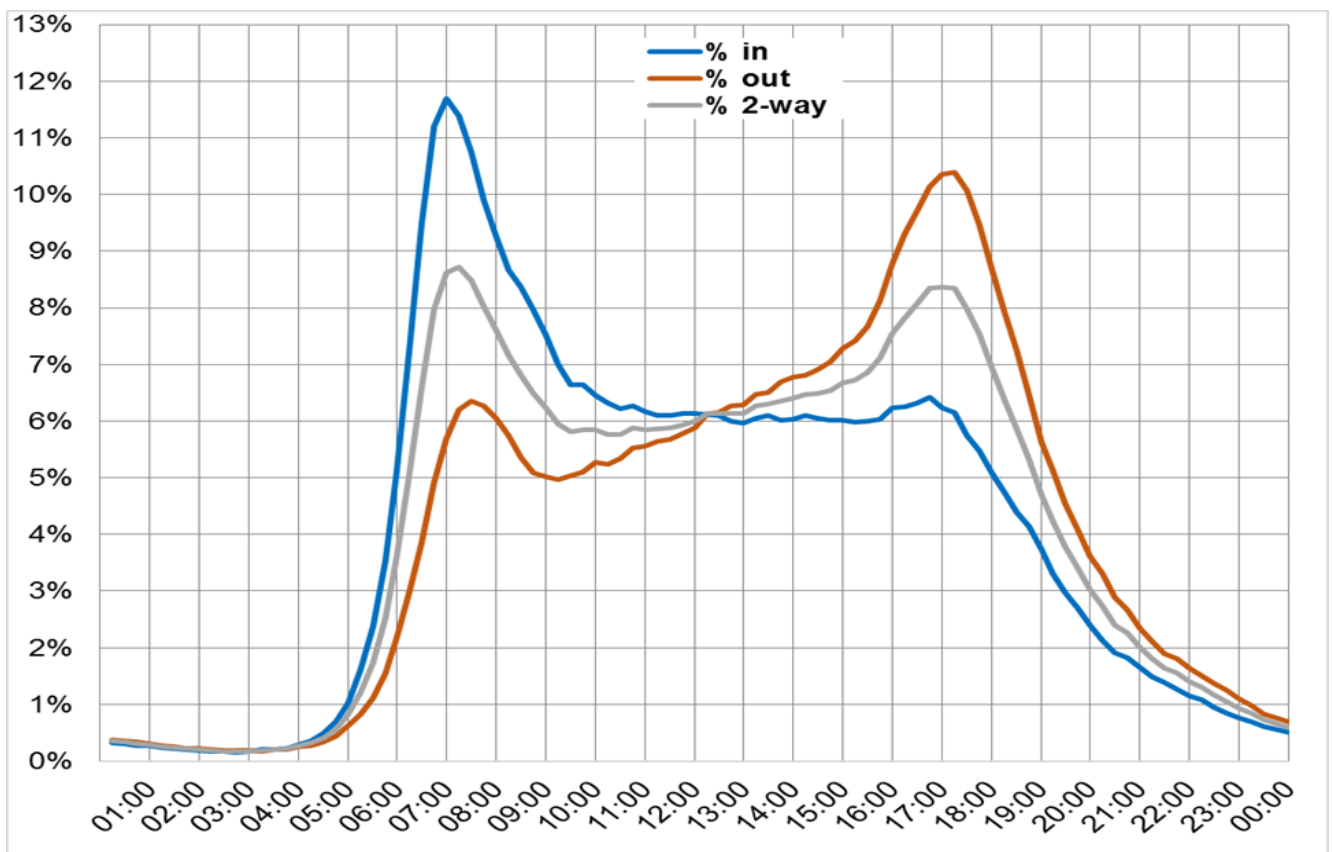
As can be seen in both Figures, the traffic volume in peak hours is substantially greater than in off-peak hours with volumes skewed "inbound" (towards work opportunities) in the morning and

“outbound” in the evening (tidal flow). During “off-peak”, taken to be daytime hours between the normal weekday peaks, flows in both directions are approximately equal.

Note too that the AM and PM peak hours each typically constitute about 8% (congested) to 10% (uncongested) of the Average Daily Traffic (ADT), while off-peak volumes during the day are around 6% of ADT.



**Figure 4: Typical Daily Volume Variation in less congested Urban Areas**



**Figure 5: Typical Daily Volume Variation in more congested Urban Areas**

Hence to get an accurate assessment of the morning, off-peak and evening peak hour volumes, it is normal to undertake traffic counts from 6:00 to 9:00, 11:00 to 14:00 and 15:00 to 18:00. Where counts are not available for all three of these periods (e.g. if count projections are taken from a transportation model or from a traffic impact study of one peak period only), it is possible to generate traffic counts in the uncounted periods based on a single peak hour volume. This is done automatically in AutoJ based on the typical weekday traffic patterns shown in Figures 4 and 5 from which the factors in section 3.4 are derived.

### **3.3. EVENT TABLE FOR TRAFFIC SIGNAL CONTROL**

When designing an intersection control, a question that needs to be answered is what period should be chosen for the intersection design? For a priority control this should be the period with the highest volume, but for a traffic signal control, any number of periods can be designed for.

In practice, it is normal for fixed-time traffic signals to have three plans to cater for the three distinct and known traffic patterns during AM, PM and off-peak periods. It is also customary to utilize “off-peak” timings at night and on weekends, as at these times the flows in both directions are also more or less equal. Alternatively, the off-peak timings can be used at night with a shorter cycle if the day off-peak plan needs a cycle time exceeding 75 seconds.

When determining the duration of the timing plan, it is standard practice to set the signal controller to introduce the peak plan about 60 minutes before the true peak hour arrives and run the plan settings for about 60 minutes past the peak. This is not only to clear peak direction traffic before and after the true peak occurs, but also to give the controller time to phase in the new programme. In an Area Traffic Control System, it takes about five to six cycles to bring all the controllers from one plan to another without having excessively long green times on an approach “waiting” for the co-ordination pulse. Hence the change takes five to ten minutes. This is also why it is generally not a good idea to detect or wait until traffic builds up before starting to change to a new plan.

Adaptive programmes such as SCOOT or SCATS move towards a new optimum in response to a traffic pattern change, but, because this takes time, often do not reach the optimum before the traffic pattern changes again. The lack of optimal timing can, in itself, be a reason why the traffic pattern might change. That is one reason why a fixed time programme will outperform an adaptive programme in a typical urban peak hour when a predictable and consistent traffic pattern exists.

Normally therefore peak plans will run from 6:00 to 9:00 in the morning and from 15:30 to 18:30 in the evening with the off-peak plan running at all other times. If volume patterns vary from the norm, e.g. around shopping centres, conference centres, sports stadiums or tourist resorts, then

more than the standard three daily plans, alternatively vehicle actuated programmes, can be considered.

Some engineers may prefer to set the signals for the “worst” case based in the Peak Hour Factor (PHF) (peak 15 minutes multiplied by four). In most urban areas the Peak Hour Factor is quite high, so it would make little difference if the signals are set for the peak 15 minutes or the actual peak hour. Also, signal plans typically run for a period of 3 hours or longer. Hence optimizing for the peak 15 minutes rather than for the peak hour may not be optimal over the whole period and both V/C and delay will be overestimated.

For full vehicle actuation (VA) the extension time needs to be determined. This is the difference between the minimum allowable green time and the green time which would be allocated to the maximum expected volume. The maximum VA cycle time with full extensions can be more than the 120 seconds recommended maximum for fixed time signals.

### 3.4. AVERAGE DAILY TRAFFIC (ADT)

The Average Daily Traffic (ADT) is taken to be the 24-hour traffic count during a typical week day in an urban area. As the full 24 hours are rarely counted at intersections, ways of estimating ADT from hourly counts have been derived.

It can be shown that the average two-way AM peak hour (6:45 to 7:45, or 7:00 to 8:00) and PM peak hour (16:30 to 17:30) is 8.9% and 8.7% of ADT respectively (standard deviation 1.2%) while the off-peak volume is 6.0% of ADT (std. dev. 0,6%).

The ADT can be derived from hourly counts. This is best done by giving approximately equal weighting to the three hourly volumes used, using factors of 3.75 (AM), 5.55 (off) and 3.85 (PM). Hence  $3.75 * 8.9 + 5.55 * 6.0 + 3.85 * 8.7 = 33\% + 33\% + 34\% = 100\%$  (Sampson 2017).

The 12-hour (6:00 to 18:00) count was found to be 82% of ADT in typical urban conditions and if available would be used in preference to hourly counts for estimating ADT.

### 3.5. WEEKLY AND MONTHLY TRAFFIC VARIATIONS

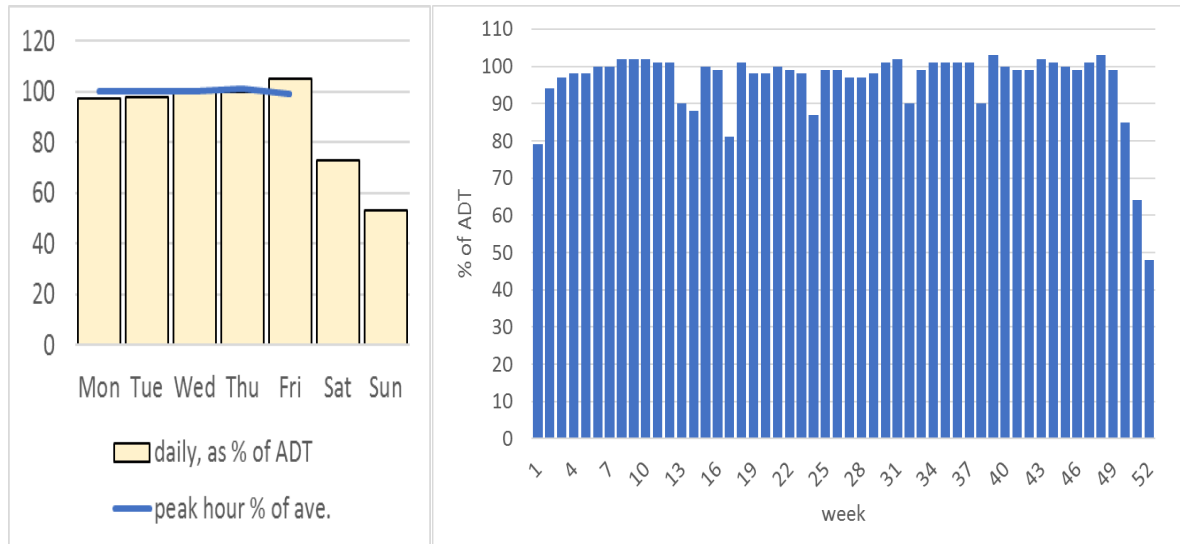
The following variation in traffic volumes by day of week and month of year were found.

**Table 2: Monday to Sunday traffic volume variations** (Sampson 1983/2017):

Day of week	Mon	Tue	Wed	Thu	Fri	Sat	Sun
peak hour % of ave.	100	100	100	101	99		
daily, as % of ADT	97	98	100	100	105	73	53

**Table 3: Monthly volume variations** (Sampson 2017):

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
%	91	104	105	97	100	100	104	101	103	103	104	87



**Figures 6 and 7: Daily and Weekly Volume Variations**

In Table 2 it is noted that while a Friday is the busiest day of the week overall, it is no more busy than other days during the peak hours. In Figure 6 the lower counts on Saturday and Sunday are clear whereas in Figure 7 the effect of school and public holidays is obvious.

It was concluded from the above that, because the volumes during peak hours are so consistent (upper row Table 2, blue line, Figure 6), a traffic count can be done on any day of the week or month of the year provided that the count is not taken within a week where there is a public or school holiday or on the Monday or Friday on either side of a long weekend.

### 3.6. LANE VOLUME VARIATIONS

In Table 4 the actual lane distribution on multi-lane highways is shown (Traffic Flow Variations in Urban Areas, Sampson 2017). Vehicles do not distribute themselves equally in each lane with light vehicles preferring the fast lanes and heavy vehicle generally using the slower lanes (but not exclusively, much to the annoyance of some car drivers).

**Table 4: Lane Usage on Multi-Lane Roads (1 is left or slow lane)**

	<b>2 lanes</b>		<b>3 lanes</b>		<b>4 lanes</b>	
<b>Lane no.</b>	<b>All</b>	<b>HV only</b>	<b>All</b>	<b>HV only</b>	<b>All</b>	<b>HV only</b>
1	42	70	24	37	18	25
2	58	30	38	45	28	42
3			38	18	30	24
4					24	9

### 3.7. EQUIVALENT VEHICLE UNITS

The number of left, straight and right turning vehicles on each approach is fundamental to any traffic analysis. Before the counted volume is used in an analysis however, it needs to be converted to the effective volume, or equivalent vehicle units (evu).

Heavy vehicles (HV) are slower and larger than passenger cars. To compensate for this, the actual counted volume should be increased. Typically, a heavy vehicle is considered equal to two light vehicles, but its effect could be three or more (e.g. finding gaps at a Stop street).

The effective volume is therefore the actual volume multiplied by:

$$1 + (F-1) * \% HV$$

where F is the weighting factor (typically 2, as above).



## 4. CAPACITY OF PRIORITY CONTROL DEVICES

### 4.1. INTRODUCTION

While many formulae have been developed in international literature, the capacity equations used for Stop streets (TWSC), roundabouts and right turning in gaps were extracted primarily from the **Highway Capacity Manual** (TRB 2000 and 2010). The equations used for all-way Stops were derived from first principles as described below.

### 4.2. TWO-WAY STOP CAPACITY

The capacity of a two-way Stop is determined by the following HCM equation:

$$C_x = V_c * e^{-q_c * t_c} / (1 - e^{-q_c * t_f})$$

$C_x$  = capacity of movement x (veh/hr);

$V_c$  = conflicting flows to which minor movement must give way (veh/hr plus ped/hr);

$q_c = V_c / 3600$  = conflicting flows (veh/sec);

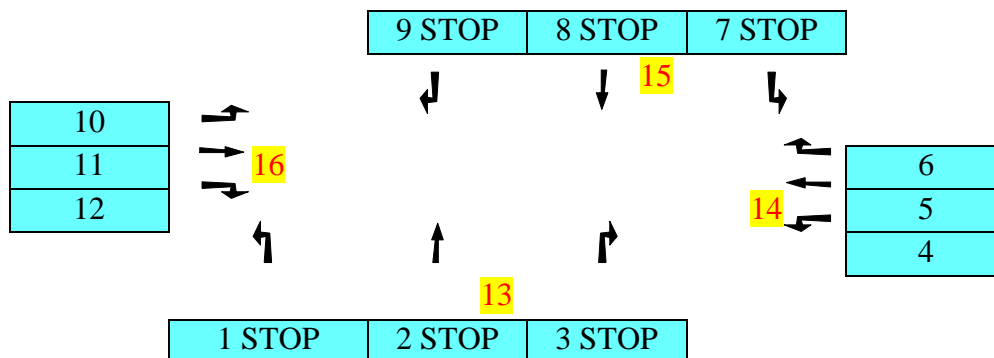
$t_c$  = critical gap (time required for the first waiting vehicle to accept a gap in traffic) (secs);

$t_f$  = follow up time or headway (i.e. the time between following vehicles taking the same gap) (secs);

$e$  = the base of natural logarithms.

#### 4.2.1. DERIVATION OF $V_c$ THE CONFLICTING FLOW

In Table 5 below, movements are numbered as per this sketch (movements 13 to 16 are pedestrians crossing in front of the approach):



The calculation of opposing flow is done using example movements '1', '2', '3' and '12' where '1', '2' and '3' stop at a Stop sign and '12' is a "free" right turn from the main road.

**Table 5: Conflicting volumes**

Movement x	$V_c$ , conflicting flow is the sum of:
'1' left turn from Stop	'4'*i + '5'/j + '13'*0.5 + '16'
'2' straight from Stop	'4'*i + '5' + '6'*2.0 + '10'*0.5 + '11' + '12'*2.0 + '13'*0.5 + '15'*0.5
'3' right from Stop	'4'*i + '5' + '6'*2.0 + '11' + '12'*2.0 + '7'*0.2 + '8' + '13'*0.5 + '14'
'12' right from main	'4' + '5' + '13'

$i = 0$  if movement '4' is in an exclusive turn lane;  $i = 0.5$  if in shared lane;

$j =$  the number of straight lanes.

The method used is identical to the HCM method except that the pedestrian volume directly in front of the stop line '13' is divided by two as in South Africa most drivers will take a gap and not wait for pedestrians unless they are directly in front of the vehicle. Also, most pedestrians will yield to a vehicle if it is seen to be taking a gap.

#### 4.2.2. DERIVATION OF TC THE CRITICAL GAP

Table 6 gives the critical gap acceptance factors from various references and adjusted based on modelling for South African traffic conditions and for use in the AutoJ simulation (Sampson 2016). Quite a wide variation can be noted, and the fact that a larger gap is needed for wider crossings is evident.

The values in Table 6 are for two-way cross roads with no median island. If a median island is present, the crossing can be done in two stages which would reduce the critical gap needed. That adjustment is described later.

The simulated 2-lane values are for a 10m wide crossing with no median.

**Table 6: Typical values of critical gaps found in references**

	HCM (TRB 2000)  2 lane	HCM (TRB 1985)  2 lane	Van As & Joubert (1993)  2 lane	Sampson various (1992, 2016)  2 lane	HCM (TRB 2000)  4 lane	HCM (TRB 1985)  4 lane	Van As & Joubert (1993)  4 lane	Sampson various (1992, 2016)  4 lane	Used for AutoJ  <b>2 lane</b>
left into r/about	4.1, 4.6			3.2				3.2	<b>4.4</b>
left at slip	4.1, 4.6	5.0		2.5, 5.4, 5.5		5.0		5.5	<b>4.8</b>
left from Stop	6.2	5.5	5.5	4.1, 5.0, 7.2	6.9	5.5	6.5	6.5	<b>6.2</b>
str. from Stop	6.5	6.0	5.9	6.1, 5.8, 7.5	6.5	6.5	6.3	6.8, 8.0	<b>6.5</b>
right from Stop	7.1	6.5	5.1	4.3, 6.3, 7.7, 8.0	7.5	7.0	5.4	5.7, 8.5	<b>7.1</b>
right from main	4.1	5.0	4.7	5.3, 5.5	4.1	5.5	4.7	4.7, 5.5, 6.0	<b>5.5</b>

In the HCM formula, the critical gap ( $t_c$ ) determines the rate at which capacity decays with increasing conflicting flows. The effect of using a higher critical gap is therefore a faster and greater reduction in capacity as conflicting volumes increase.

Many references (e.g. Joubert 2010) and extensive testing done for AutoJ (Sampson 2016) indicate that using the critical gap values in HCM resulted in an overestimation of capacity of the movements taking gaps, particularly as conflicting flows neared saturation.

It was found that one way to overcome this without getting negative or zero capacity was to multiply the critical gap values by a correction factor of between 1.0 and 2.0. The correction factor has a minor influence on capacity at low conflicting flows but substantially reduces capacity with mid to high conflicting flows. Although in earlier versions of the AutoJ program this correction factor was used, to be more in line with international practice it was decided to

rather use a value of critical gap on the high side (as per HCM 2000 in most instances) and to adjust this by factors developed below to take roadway crossing width and the presence or absence of shelter (median island width) into account. In addition, in AutoJ the calculated capacity is reduced by 80 vehicles per hour which is the estimated extent of the HCM overestimation at high conflicting volumes.

It can also be mentioned that motorists on the main road giving courtesy gaps to side road traffic does increase side road capacity.

The adjustments to critical gap values have been derived as follows:

1. A **crossing width** adjustment has been made to account for the longer critical gap needed on wider roads. In Table 6, some authors have found longer critical gaps are needed for crossing four lane roads. This has been translated into an adjustment for wide road crossings based on the crossing width in metres.

The standard crossing width is set at 10m (from stop line to clearance on far side) and an additional 0.02 seconds is added for each additional metre to be crossed by straight across traffic and 0.04 seconds for right turn traffic. An additional crossing time of 0.02 seconds per metre is also added for right turns from the main road. As an example, the critical gap for a straight crossing would be 0.2 seconds more for a 20-metre crossing than for a 10m crossing.

The values of 0.02 and 0.04 were obtained firstly by reference to Table 6 and then by simulating various widths of crossings and noting their effect on capacity. Further research is needed to refine these values.

2. **Median islands** enable a two-stage crossing, provided the median is wide enough, but even a narrow median provides some shelter to a crossing vehicle. For the purposes of AutoJ, the critical gap acceptance time was reduced by 0.2 seconds for every metre of median island width (based on simulation). Field studies are needed to refine this value, another opportunity for future research.
3. Adjustments are made for **heavy vehicles**. For this adjustment, every heavy vehicle is considered equal to two light vehicles, but AutoJ users can modify this figure.
4. A further adjustment needs to be made for **grade**. For each 1% upgrade, 1% was subtracted from the capacity (or added for downgrades). The latter is the same adjustment made in the HCM (TRB 2000), although according to more recent research (Bruwer MM, Bester CJ, Viljoen ES, **The Influence of Gradient on Saturation Flow Rate at Signalized Intersections**, Journal of the South African Institution of Civil

Engineering, June 2019), this should be 3% for every 1% grade change for both uphill and downhill grades.

#### 4.2.3. DERIVATION OF TF THE FOLLOW UP TIME

The **follow up time**, or headway, between vehicles following the initial gap taker has a significant influence on capacity. For example, a two second headway results in a saturation flow of 1800 veh/hr while a three second headway reduces this to 1200 veh/hr.

The HCM (TRB 2000) follow up times were compared with saturation flows ( $S$ ) reported in other literature ( $S = 3600 / t_f$ ) and after simulation, the values listed in Table 7 were found to be best.

**Table 7: The follow up (headway) times recommended**

	HCM (TRB 2000) (Kittelson) (Joubert 2010)	Used for <b>AutoJ</b> (Sampson 2016)	<i>equivalent saturation flow</i>
free flow, left		<b>1.89</b>	<i>1900</i>
free flow, straight		<b>1.80</b>	<i>2000</i>
free flow, right		<b>1.98</b>	<i>1820</i>
gaps signal, right	2.2	<b>2.00</b>	<i>1800</i>
slip, left		<b>2.50</b>	<i>1440</i>
mini-circle, left		<b>2.25</b>	<i>1600</i>
roundabout, left	2.6 – 3.1 2.1 – 2.7	<b>2.50</b>	<i>1440</i>
Stop, left	3.3	<b>3.30</b>	<i>1090</i>
Stop, straight	4.0	<b>4.00</b>	<i>900</i>
Stop, right	3.5	<b>3.50</b>	<i>1030</i>
all-way Stop, left		<b>2.58</b>	<i>1394</i>
all-way Stop, straight		<b>3.16</b>	<i>1140</i>
all-way Stop, right		<b>3.32</b>	<i>1085</i>

The all-way Stop saturation flow values in Table 7 were derived from first principles based on the fact that vehicles at an all-way Stop do not take gaps but operate on a first-come first-served basis,

which, in the absence of conflicting traffic, is how quickly a driver can stop, look and proceed. Because gaps do not have to be judged, the capacity at an all-way Stop is higher than at a conventional Stop.

#### 4.2.4. MULTIPLE APPROACH LANES

For multiple approach lanes, each additional lane does not have the same saturation flow as a single lane. There are two main factors to consider:

- Vehicles do not distribute themselves exactly equally in each available lane even if they are the same movement.

Straight through vehicles share the lanes approximately equally, but where there is more than one turning lane, vehicles will stagger themselves rather than turning together. At Stop streets, the middle lanes are avoided as visibility in both directions is limited. Site observations suggest that each additional lane at a Stop line is equivalent to 0.6 of a full lane. This capacity reduction for additional lanes is quite severe and does require further research; however more than two lanes at Stop streets are quite rare and should be avoided anyway.

- Shared lanes are avoided in preference to exclusive lanes, e.g. straight through vehicles will avoid being stuck behind a right turner in the same lane if possible. The recommended adjustment for a shared left and straight lane is a 4% reduction in capacity; for shared right and straight, a 10% reduction; and for a shared left, right and straight, an 8% reduction. Although similar factors are mentioned in references they also have not been well researched.

#### 4.2.5. STOP STREET CAPACITY CALCULATION SUMMARY

In summary, the Highway Capacity Manual formula for Two-Way Stop Capacity was found to be the best available but it is necessary to apply factors to adjust the critical gap and headway to account for crossing width, number of lanes, two stage crossings, heavy vehicles, grade and the underestimation of the effect of conflicting flows. Adjustments are also suggested to the HCM method to determine opposing flow volumes to allow for pedestrians who tend to cross behind or yield to the first vehicle on the stop line (0.5 of the pedestrian crossing volume is used whereas HCM uses full volume).

### 4.3. ALL-WAY STOP CAPACITY

As no suitable method to automatically calculate all-way Stop capacity was found in the references consulted (other than cumbersome iteration methods), the all-way Stop capacity for AutoJ has been developed from first principles.

Motorists at an all-way Stop do not take gaps but make their decision to proceed on a first-come first-served basis. In all cases it is assumed motorists behave legally, bring the vehicle to a complete stop, wait for their turn before proceeding and do not “follow on” the vehicle in front when it is not their turn. While this does not always happen in practice, for a fair comparison with other controls it was decided not to account for illegal behaviour.

The two possible extremes of opposing flow facing a motorist at the stop line are 1) no conflicting vehicles (allowing maximum stop line flow) and 2) saturation conflict (minimum stop line flow):

1. The **maximum possible capacity** per lane (saturation flow) occurs when a queue of vehicles arrives at the stop line and there are no other vehicles or pedestrians on any approach. The saturation flow rate is therefore determined only by the headway between following vehicles. From Table 7 (above), the saturation flow rate of an unopposed left, straight and right turn from a Stop can be calculated to be 1 394, 1 140 and 1 085 veh/hr respectively before adjusting for heavy vehicles and grade.
2. The **minimum capacity** occurs when a vehicle arrives at the stop line and there are vehicles and pedestrians on every other approach that have arrived before the subject vehicle. This vehicle will have to wait for the vehicles and pedestrians proceeding straight from the approaches to the left and right (assumed to go together), the vehicles turning right from the approaches to the left and right (also assumed to turn together), and the right turner opposite. Other opposing pedestrians are assumed to cross while the vehicle waits for its turn. The total wait will therefore be the combined time for the first vehicle on each of the opposing movements to clear.

The time,  $t$ , for a movement to clear is calculated from the formula  $s = u t + 0.5 f t^2$ , where  $s$  is the clearance distance (from stop line to far side),  $u$  is the initial velocity (zero from a stop),  $f$  is the acceleration rate (considered to be  $2.0 \text{ m/s}^2$  for normal car ready and anxious to take their turn and allowing for the fact that the vehicle does not always have to clear completely before the next vehicle begins to move), hence time to cross  $t = s^{0.5}$ ; a reaction time is not added as the vehicle has been waiting.

As an example, if the intersection is 9 metres wide, each vehicle would need 3.0 seconds to clear and the minimum capacity of each approach would be  $3600 / (4 * 3.0) = 300$  vehicles per hour whereas a 16m crossing will have a capacity of 225 veh / hr. Because left turners can also turn with side road right turners they get two opportunities to turn, therefore their

minimum capacity is double the straight and right capacity (provided they have an exclusive left turn lane and are not blocked by other movements).

The minimum capacity of vehicles at the stop line at an all-way Stop in a saturation flow situation is substantially higher than at a two-way Stop as vehicles do not have to find gaps but simply take their turn regardless of the other approach demand volumes.

3. Having established the maximum and minimum capacities, the **actual capacity** under varying traffic volumes needs to be calculated, obviously somewhere between minimum and maximum. This was originally done in AutoJ by calculating the “unused” capacity on each approach and adding that to the minimum available capacity. For example, if the volume on an approach in the above example was 100 vehicles per hour and the capacity was 300, this would “release” a capacity of  $300 - 100 = 200$  vehicles per hour to be used by the other movements. This released capacity is shared by the other movements up to the maximum capacity.

In later versions of AutoJ the V/C of the intersection as a whole is calculated. If the V/C is 1.0 or above, the minimum capacity applies. If the V/C is 0.0, the maximum applies. If the V/C is anywhere in between, the difference between maximum and minimum capacity multiplied by  $(1 - V/C)$  is added to the minimum capacity for each movement.

In multi-lane situations, the same principles apply but the volume per lane, not the total volume, is used to determine if there is spare capacity.

These are the calculations built into AutoJ.

As an aside, some users have noted that while the program determines that substantial delays should occur at an all-way Stop intersection, in practice it is observed to flow with little delay. The reason for this is that most vehicles observed do not stop at the stop line and that many “follow on” movements occur without stopping as well. As stated above, the author decided not to adjust the calculations for such illegal practices. What is recommended in these situations is that the all-way Stop control be changed to a mini-circle, which would make what is observed to happen in practice at the intersection (yielding and following on) legal and safe.



#### 4.4. ROUNDABOUT CAPACITY

Six roundabout capacity equations were tested, namely Tanner 1962, McDonald and Armitage 1974 and 1978 (Van As & Joubert 1993), Highway Capacity Manual (TRB 2000), Highway Capacity Manual (TRB 2010) and Kittelson and Associates (Rodegerdts 2015).

It was noted that although in the HCM (TRB 2000) the equation for roundabouts appears to be quite different from the equation for two-way Stops, it is in fact the same equation with different symbols. Also, for a roundabout HCM 2000 specifies the number of lanes may not exceed one and the conflicting flow may not exceed 1 200 vph.

Based on later research, although the HCM 2000 formula below was adopted for AutoJ, the factors and limits were adjusted as per HCM 2010 and Kittelson (Rodegerdts 2015).

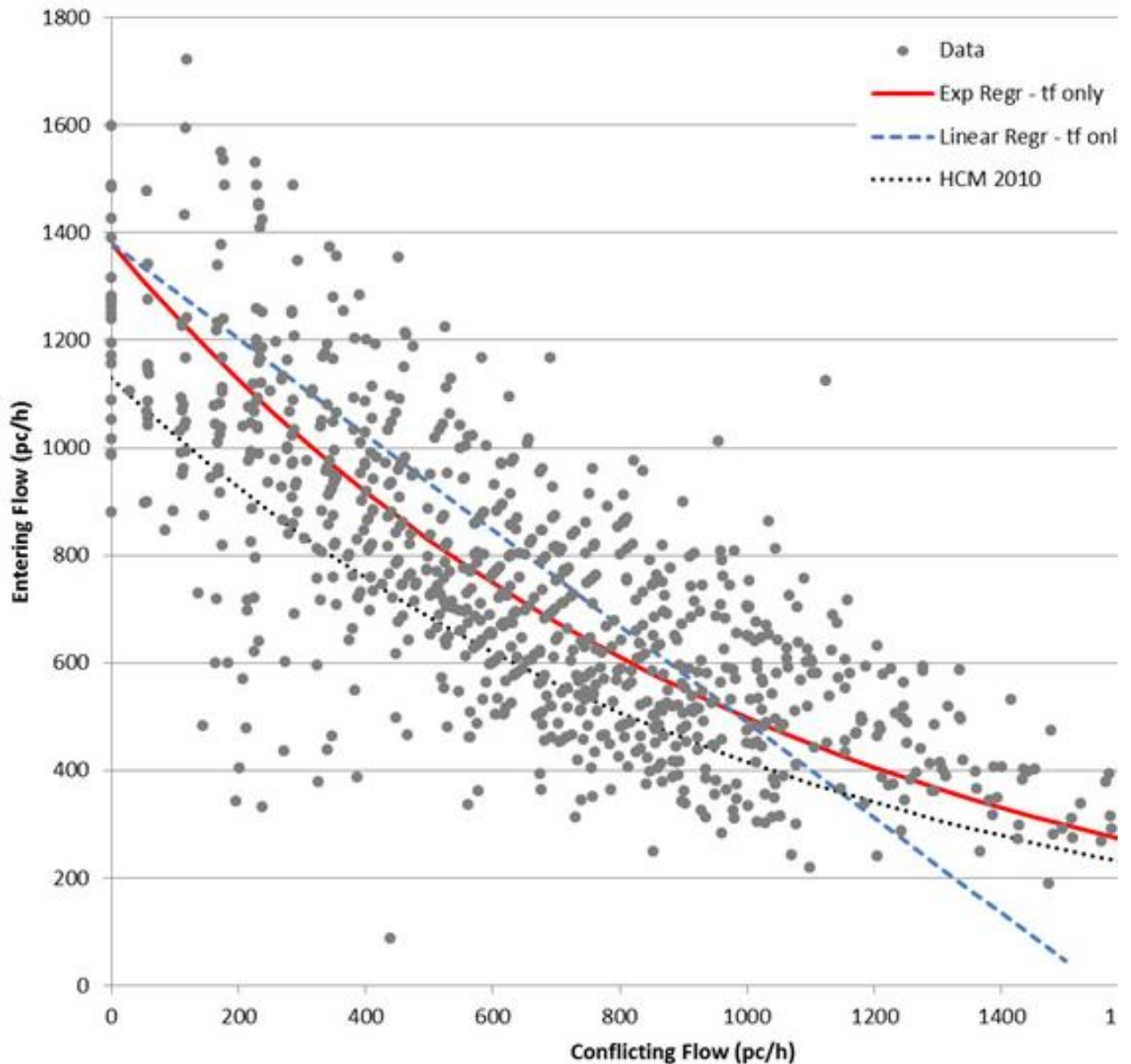
The formulae from the respective researchers are:

1962 Tanner	$C = Q * (1 - t_z * q) / (1 - e^{-q * t_f}) * e^{-(t_c - t_z) * q}$
1974 McDonald + Armitage	$C = N * Q / (e^{q * t_c} - 1)$
1978 McDonald + Armitage	$C = S * (1 - t_z * q) * e^{-(T - t_z) * q}$
2000 HCM	$C = Q * e^{(-q * t_c)} / (1 - e^{(-q * t_f)})$
2010 HCM	$C = S * e^{(-B * Q)}$
2015 Kittelson	$C = S * e^{(-B * Q)}$

Where:

C =	capacity
q =	conflicting volume (veh / sec)
Q =	conflicting volume (veh / hour)
S =	saturation flow (veh/hr) (1 656 M+A 1978, 1 130 HCM 2010, 1 420 Kitt 2015)
t <sub>c</sub> =	critical gap (secs)
t <sub>f</sub> =	following gap / headway (secs)
t <sub>z</sub> =	min headway for circulating vehicles (sec)
T =	lost time
N =	number of circulatory lanes (one)
B =	factor (-0.001 HCM2010, -0.00085 Kitt2015)

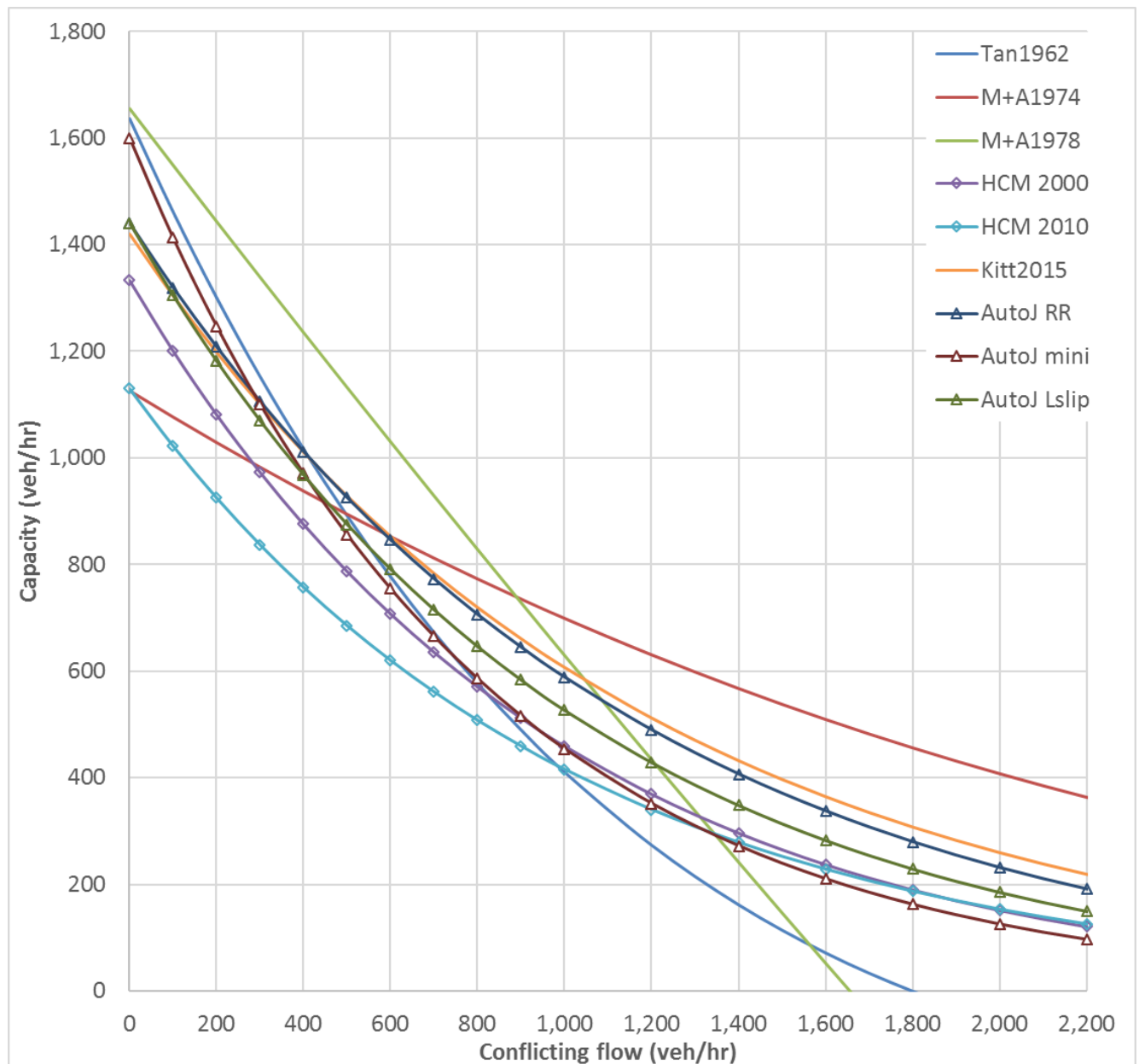
In 2012 and 2013, Kittelson and Associates collected data from 23 roundabout sites throughout the USA (NCHRP Report 572 data: 2013). That data together with fitted curves is reproduced in Figure 8 (Rodegerdts 2015).



**Figure 8: Field data with HCM 2010 formula compared to Linear and Exponential regression (Rodegerdts 2015)**

It can be noted that the HCM 2010 formulation appears to underestimate roundabout capacity (compare the HCM curve with the exponential regression curve).

The six formulae described above with their default values were then plotted to the same scale and the AutoJ formulation for a Roundabout, mini-circle and left turn slip was added to provide Figure 9. The HCM 2010 formulation is common to Figure 8 and Figure 9.



**Figure 9: Roundabout capacity determined by different authors**

The Tanner formula at some point gives negative capacity at high conflicting flows (in this example exceeding 1 800 vph) and therefore cannot be considered for use in a simulation program. The two M + A formulae had the same problem but also did not fit the field data well. The HCM 2010 formula clearly underestimated the average roundabout capacity as did HCM 2000 with the default factors. The Kittelson formulation was clearly the best fit of the field data (Figure 8).

It was therefore decided to use the HCM (TRB 2000) formula for roundabouts with the factors in Tables 6 and 7 which gives almost exactly the same result as Kitt2015 (Rodegerdts 2015) as can be seen in Figure 9.

The equation used for AutoJ was therefore:

$$C = V_c * e^{-qc * t_c} / (1 - e^{-qc * t_f})$$

with the critical gap  $t_c$  set at 4.4 seconds (Table 6) and the follow up gap  $t_f$  set at 2.5 seconds (Table 7). This formula gives the AutoJ RR graph in Figure 9.

#### 4.5. MINI-CIRCLE CAPACITY

At low conflicting flows, mini-circles operate almost like roundabouts. However, as flows increase the circle is too small to operate on modern roundabout principles (i.e. gap acceptance) and starts to operate on a first come – first served basis (i.e. like an all-way Stop).

To simulate this, it was found decided to use the same roundabout formula but to set the critical gap and follow up gap to 5.6 and 2.25 seconds respectively. The effect is shown in Figure 9.

It is also interesting to note that because of the widespread illegal ignoring of the Stop at an all-way Stop, it too operates in a similar manner to a mini-circle at low volumes, but not as well as there will always be those law-abiding citizens who do stop despite their fellow drivers behaviour.

#### 4.6. PEDESTRIAN CAPACITY

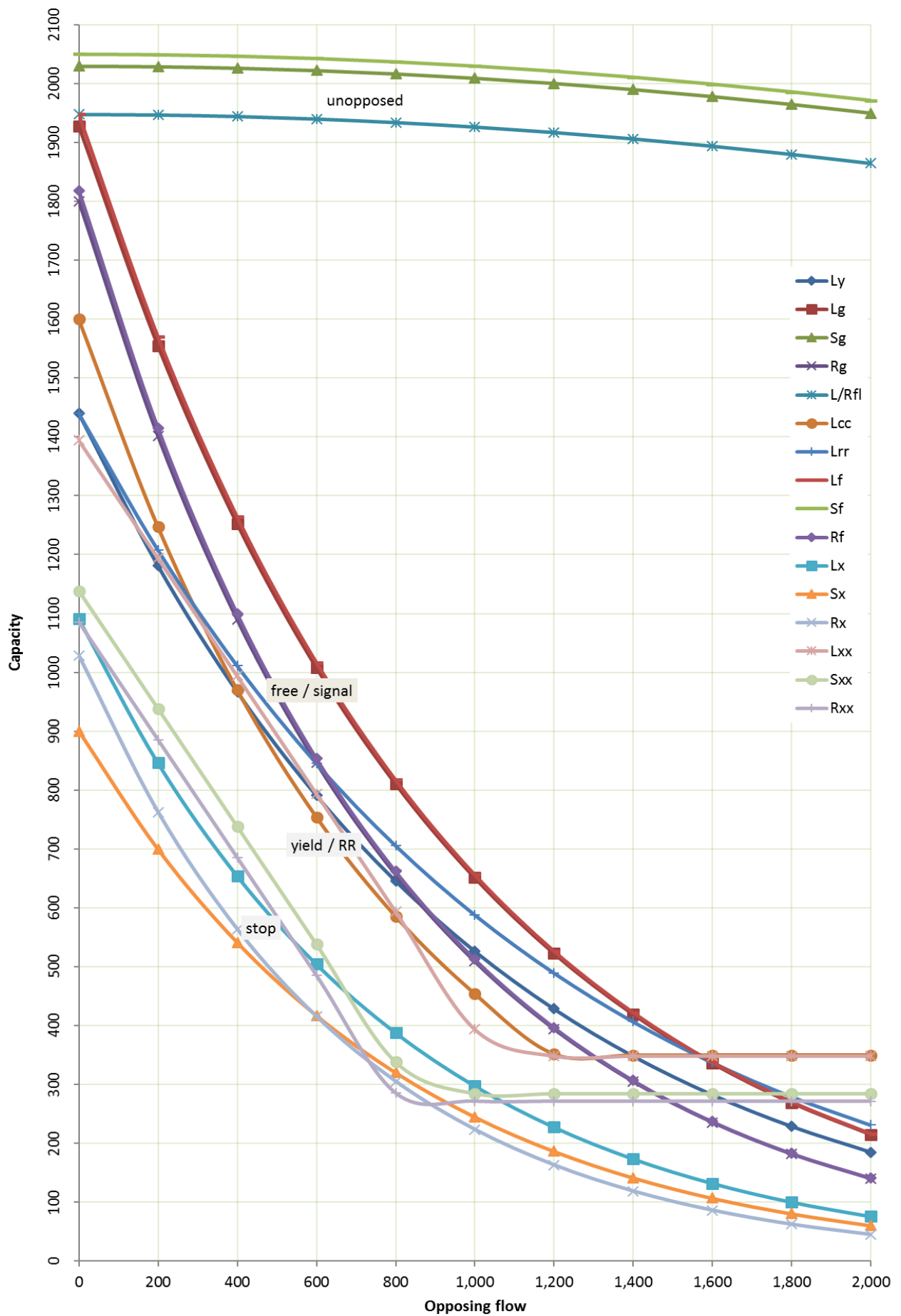
The saturation flow rate of pedestrians is 4 800 pedestrians per hour per metre crossing width (TRB 2010). This would apply to pedestrians with right of way such as crossing in front of vehicles at a Stop or Yield sign or at a green traffic signal.

For pedestrians crossing uncontrolled roadways this cannot be achieved. In these circumstances it is assumed pedestrians will take the same gaps as vehicles. The pedestrian capacity is therefore taken to be the saturation flow rate for pedestrians multiplied by vehicle capacity to saturation flow ratio at a Stop street.

#### 4.7. RESULTS

The results of applying the adopted formulae (as modified) on capacity per lane are summarized in Figure 10.

In the figure  $L_y$  is left at a yield sign;  $L_g$ ,  $S_g$ ,  $R_g$  and  $R_{fl}$  are left, straight, right and right flash at a signal with 100% green ( $L_g$  must yield to pedestrians and  $R_g$  to pedestrians and opposing traffic while  $S_g$  and  $R_{fl}$  are unopposed, influenced only by surrounding vehicles);  $L_{rr}$  is left turn into roundabout;  $L_f$ ,  $S_f$  and  $R_f$  are left, straight and right free flows (no control, but turners must take gaps in vehicular and pedestrian traffic);  $L_x$ ,  $S_x$  and  $R_x$  is left, straight and right after Stop;  $L_{xx}$  and  $SR_{xx}$  are left, straight and right at an all-way Stop.



**Figure 10: Effect of conflicting vehicle and pedestrian flows on capacity**

#### **4.8. CONCLUSION ON CAPACITY OF PRIORITY CONTROLS**

From an examination of international literature, it has been found that the Highway Capacity Manual formulae are appropriate to use for Stop and Roundabout priority intersection capacity calculations. However, it was also found that it is necessary to expand and refine the gap acceptance and follow up values to cater for wider intersections, intersections with median islands and intersections with opposing volumes higher than the limits determined in HCM.

Furthermore, it was found that no suitable analysis method for all-way Stops was available in HCM or other literature (other than an iterative method which is not practical) and a new formula, calculated from first principles, is derived for these situations.

## 5. CAPACITY OF TRAFFIC SIGNALS

### 5.1. INTRODUCTION

This section considers the Capacity of Traffic Signalized Intersections.

### 5.2. CAPACITY CONSIDERATIONS

Traffic signal capacity is nominally  $C = S * (g/c)$  where  $S$  is the saturation flow,  $g$  is the green time and  $c$  is the cycle time.

However, the calculation of capacity of traffic signals must also adjusted for geometric considerations (e.g. grade, number and width of lanes, turning penalties), opposing vehicular and pedestrian flows, start-up lost time and inter-green overruns. Vehicle composition, including heavy vehicles and buses, are accounted for using the effective volume, or evu (equivalent vehicle units), calculations described in Chapter 3. The derivation of each of the remainder of these components is described below.

#### 5.2.1. SATURATION FLOW RATE

The recommended saturation flow rate value for an “infinite” queue of light vehicles flowing freely in a single lane in ideal conditions on a flat grade proceeding straight ahead is 2 000 vehicles per hour, although higher and lower flow rates have been measured on occasions.

In a paper **Saturation Flow Rates** by C J Bester and W L Meyers, University of Stellenbosch, July 2007, the following previous studies were quoted:

**Table 8: Previous studies’ saturation flow rates**

Study	Country	Mean veh/hr/lane
Kimber et al	UK	2080
H E L Athens	Greece	1972
Hussain	Malaysia	1945
Bonneson et al	USA (Texas)	1905
Webster & Cobbe	UK	1800
Branston	UK	1778
Miller	Australia	1710
De Andrade	Brazil	1660

Shoukry & Huizayyin	Egypt	1617
Coeyman & Meely	Chile	1603
Bhattacharya & Bhattacharya	India	1232

As part of the Bester and Meyers study, roads in the Western Cape were also measured. The minimum flow rate for straight through movements measured was 1553 with a maximum 2605 veh/hr/lane. A base rate capacity in a 60km/h speed limit zone of 2076 veh/hr was suggested as the best fit, with a right turn saturation flow rate of between 1840 and 1920 veh/hr/lane (C J Bester and W L Meyers, 2007).

In most **cities** and metropolitan areas in South Africa, drivers are aggressive, follow closely, take small gaps and generally maximize capacity. The saturation flow is expected to easily reach a flow rate of 2 000 vehicles per hour per lane in these conditions,

In **towns**, drivers are generally more relaxed and take longer to take a gap. They also follow less closely. In towns therefore, 1 800 vehicles per hour per lane is suggested as a more appropriate saturation flow to use (with all gap acceptance and headway adjustments made accordingly). This saturation flow rate of 1 800 veh/hr is commonly used in traffic studies.

In villages and rural areas, a further adjustment to 1 600 vehicles per hour per lane could be applied.

These adjustments are however estimates by the author based on the research above as well as other references (Sampson 1992, 2016).

#### 5.2.2. TURN CAUTION

The saturation flow rate is reduced by 5% due to increased headways caused by slowing down and caution while turning left or right in a single lane. Turning saturation flow is therefore  $2\,000 \times 0.95 = 1\,900$  vehicles per hour.

#### 5.2.3. GRADE

For each 1% up-grade, the saturation flow is reduced by 1% and for each 1% down-grade, the saturation flow is increased by 1%, as per HCM (TRB 2000).

#### 5.2.4. PEDESTRIAN INTERFERENCE

On the steady green disc traffic light indication, turning vehicles must yield to pedestrians using the crossing. As the effect of a pedestrian is much the same as an opposing vehicle in these



circumstances, the number of opposing pedestrians are added to the number of opposing vehicles for purposes of calculating turning vehicle capacity.

Straight through vehicles, and vehicles using a flashing green arrow, are however not affected by pedestrians as they have exclusive right of way.

#### 5.2.5. PEDESTRIAN CAPACITY

The saturation flow rate of 4 800 pedestrians per hour per metre crossing width is converted to capacity by multiplying by the pedestrian “**green man**” time divided by the cycle time, per metre of crosswalk (most pedestrian crosswalks are 3 m wide). If pedestrian heads are not present, these values are nevertheless set to what the green man time would be, i.e. the stage length less the pedestrian clearance time.

#### 5.2.6. RIGHT TURNS IN GAPS AT SIGNALS

In gap acceptance situations, the saturation flow of right turners is severely affected by opposing vehicular and pedestrian traffic. The effect is worsened at traffic signals because the opposing traffic is concentrated to only being able to proceed when the traffic signal is green. Hence the equivalent opposing flow rate is much higher than it would be without a signal.

To equate for this, the effective opposing volume must be increased by dividing by the green to cycle time ratio. The saturation flow for right turning vehicles taking gaps can then be calculated by using the same gap acceptance formulae as described for “free flow” priority intersections (Chapter 4).

Therefore, the saturation flow rate of right turners taking gaps can only be determined when the green time is known. However, the green time allocated to each movement depends on the movement’s capacity and that the capacity is not known until the green time is calculated. Hence an initial estimate of the green to cycle ratio for two, three and four stage signals is required.

This is done in AutoJ (Sampson 2016) using the planning method described in the US Department of Transportation Traffic Control Systems Handbook (June 1976) where the volume per straight and left turn lane is taken at unity but the volume in a right turn lane is doubled for signal timing purposes.

#### 5.2.7. RIGHT TURNS DURING INTER-GREEN

In addition to taking gaps, right turning vehicles can also clear during the inter-green period. Traditionally this is accounted for by allowing two right turners per cycle. In AutoJ (Sampson 2016) this is refined by considering the storage space within the intersection for right turners.

Wider and multi-lane intersections allow more vehicles to store and turn during inter-green. The formula used is:

$$N_i = (W/15 + 1) * n_c * n_l$$

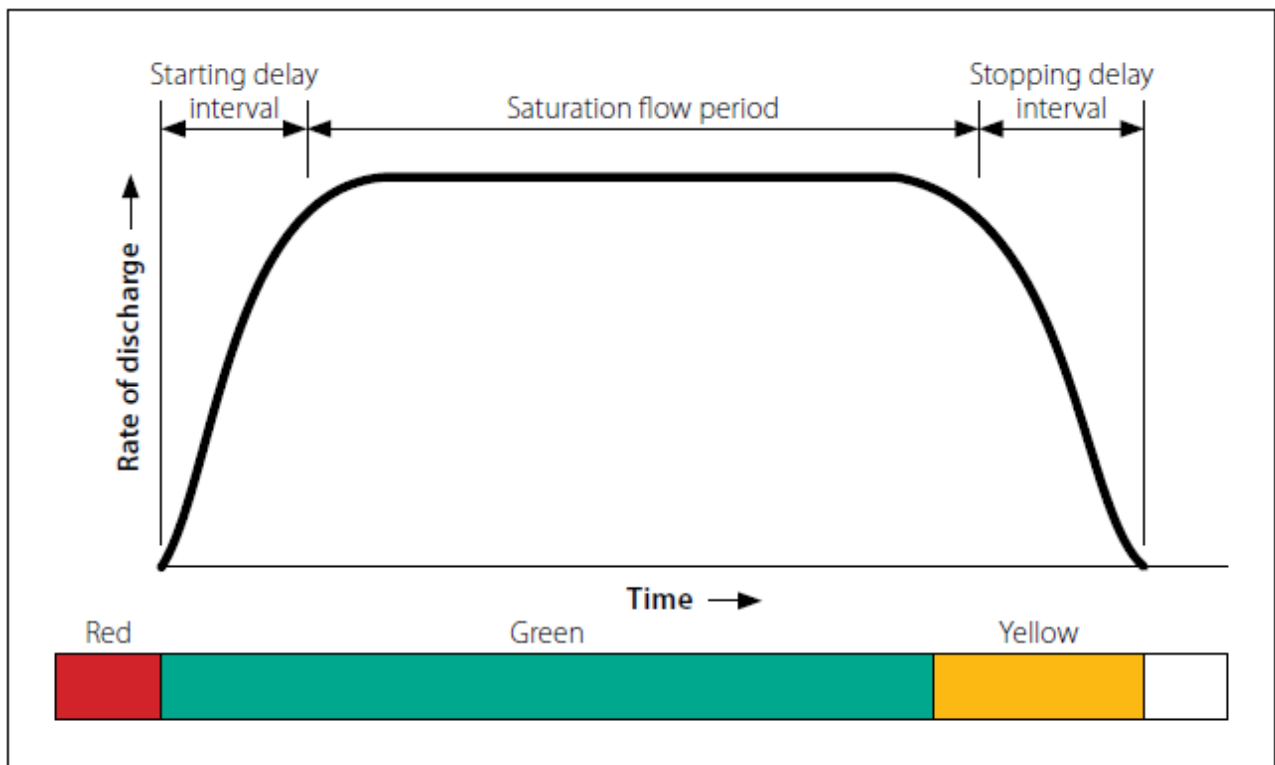
where  $N_i$  is the number of vehicles turning during the inter-green,  $W$  is the full intersection width,  $n_c$  is the number of cycles per hour (3600 / cycle time) and  $n_l$  is the number of lanes.

This works out to be 2 veh / lane / cycle if the intersection width is 15m. A 30m wide intersection would provide for 3 veh / lane / cycle turning on the inter-green, etc.

#### 5.2.8. EFFECTIVE GREEN TIME

The saturation flow is for a 100% green situation. The actual capacity is calculated by multiplying the saturation flow by the **effective** green to cycle time ratio.

When a traffic signal turns green there is start-up lost time before the saturation flow rate is reached while at the end of the green phase there is inter-green overrun time gained where some vehicles continue to flow on the yellow signal (Figure 11). In most studies, including this one, the starting delay and the stopping delay are considered equal and hence the **effective** green time exactly equals the **actual** green time.



**Figure 11: Discharge flow pattern across the stop line of a traffic signal illustrating start and stop delays**

## 6. DELAY AND QUEUES

### 6.1. INTRODUCTION

This Chapter considers Delay and Queues at Intersections. As with capacity formulae, the best theory and best international practices were examined. These theories and formulae were extensively tested against each other and again it was found necessary to make modifications in some instances. These are described in detail in the text below.

By definition, the delay caused by the Intersection Control Device itself is known as ***uniform***, deterministic or control delay, the delay incurred regardless of any other traffic.

The delay caused by the presence of other vehicles and pedestrians is called ***random*** or stochastic delay.

The additional delay that results when demand exceeds capacity is called ***overflow*** delay.

Combined, they form the ***total*** delay or system delay.

### 6.2. SYMBOLS

Symbols used are:

- $C$  = capacity, in vehicles per hour
- $C_s$  = practical capacity =  $0.975 * C = y * C$
- $c$  = cycle time =  $g + r$
- $g$  = effective green time =  $c * (1 - r/c)$  = actual green time; with start-up lost time found to be equal to inter-green overflow time, effective green time equals actual green time
- $g/c$  = green to cycle time, shown as  $\lambda$  “lambda” in some formulae
- $i$  = inter-green time = yellow plus all-red time
- $q$  = volume or flow, in vehicles per second =  $V / 3600$
- $r$  = effective “red” time =  $c - g = c * (1 - g/c)$  = actual red plus inter-green time
- $S$  = saturation flow in vehicles per hour
- $V$  = volume, in vehicles per hour
- $V/C$  = volume to capacity ratio
- $x$  = degree of saturation =  $V/C = q / (S/3600 * g/c)$
- $y$  = degree of saturation at practical capacity, taken to be =  $0.975 = C_s / C$

Delay terms are:

- $d_u$  = uniform delay, in seconds per vehicle = deterministic delay
- $d_r$  = random delay, in seconds per vehicle = stochastic delay

- $d_o$  = overflow delay, in seconds per vehicle, where  $x > y$
- $d$  = average total delay, in seconds per vehicle =  $d_u + d_r + d_o$
- $D$  = total delay, in vehicle-hours per hour =  $d * V / 3600$
- $Q$  = average queue length, in vehicles =  $D$  if slowing and accelerating delay is ignored
- $Q_o$  = overflow queue
- $\mu$  = co-ordination factor, 100% for perfect co-ordination (all vehicles arrive during green phase), 50% for random arrivals, 0% for all vehicles arriving during the red signal.

### 6.3. DELAY AT PRIORITY JUNCTIONS

The delay equations given in different sections of the Highway Capacity Manual (TRB 2000 and 2010) for a two-way Stop, all-way Stop and roundabouts appear to be different, but with nomenclature adjustment and with  $T$ , the analysis time, equal to 1 (one hour) all are the same:

$$d = 3600 / C + 900T * (x - 1 + ((x - 1)^2 + 3600 / C * x / 450)^{0.5}) + 5$$

According to HCM, if the degree of saturation  $x$  exceeds 0.9, the analysis period must be lengthened from the recommended 15 minutes to include the full period of oversaturation. This implies extending the analysis period until  $x$  again drops below 0.9. No provision is therefore made for oversaturated delay.

For AutoJ it was decided to test this equation for  $x < 0.9$ , extend it to  $x < 1$  and make the same modification for oversaturated conditions as was made for traffic signals (with the under-saturated factor of  $(x - 1) = (1 - 1) = 0$  when saturation is reached), resulting in the following equation for  $x \geq 1.0$ .

$$d = 3600 / C + 900 * (3600 / C * x / 450)^{0.5} + 5 + 1800 * (1 - 1 / x)$$

The “+ 5” seconds in the equations above allow for deceleration and acceleration delay required by the forced stop at the Stop street. For a yield sign, a full stop is not necessary but there will be some delay slowing for the yield, even if no conflicting vehicles are present. In this case the “+5” seconds is reduced to “+2” seconds.

For the “free” right turn from main road case the same delay equation is used, but the right turn vehicle does not have to stop if a gap is available. If a gap is not available, the delay is accounted for in the delay equation. The “+5” seconds is therefore not added in this case (i.e. it is set to zero).

The results of the above analyses and applying the derived formulae are summarized in Figures 12a and 12b.

#### 6.4. DELAY AT UNDER-SATURATED TRAFFIC SIGNALS

The basis for the formula for signal delay was extracted from the longest standing authority on the subject, Webster and Cobbe (1966), **Traffic Signals**. The formulae from several other references quoted in van As and Joubert (1993), **Traffic Flow Theory**, were also tested, including Miller 1968 and TRANSYT, plus the BPR, Davidson, JHK and HCM formulae described in the text below.

The Webster and Cobbe (1996) (W+C) formula for delay is:

$d = c * A + B / q - C$ , with

$$A = (1 - g/c)^2 / (2 * (1 - g/c * x))$$

$$B = x^2 / (2 * (1 - x))$$

$$C = 0.65 * (c / q^2)^{1/3} * x^{(2 + 5 * g/c)}$$

W+C concluded a rough approximation of delay is given by

$$d = 0.9 * (c * A + B / q),$$

where  $c * A = d_u$  and  $B / q = d_r$ .

Delay tends to infinity as  $x$  tends to 1; hence W+C formula is limited to  $x$  less than 0.975.

##### 6.4.1. UNIFORM DELAY

The first part of the W+C equation (A) simulates uniform delay, the delay caused by the control device itself, irrespective of any other traffic using the intersection. However, upon testing the formula it was found that it only applied to situations where 50% of the traffic arrived during the green interval. Modifications to the W+C equation are therefore necessary to cater for situations where less or more than half of the traffic arrives during the green interval.

In AutoJ (Sampson 2016) a co-ordination factor  $\mu$  has been introduced (directly related to the Highway Capacity Manual (TRB 2000) arrival types). The value of  $\mu$  ranges from 100%, for perfect co-ordination with 100% of vehicles arriving during the green signal, to 0% with all vehicles arriving during the red period. Random or uniform arrivals would approximately equate to a 50% co-ordination value.

To incorporate the co-ordination effect therefore, the divisor of 2 in the W+C equation A was replaced by the factor  $(1 - \mu)$  in the multiplier. Hence:

$$A = (1 - \mu) * (1 - g/c)^2 / (1 - g/c * x)$$

It will be noted that with random arrivals ( $\mu = 50\%$ ), the formula is identical to the W+C formula A, but with perfect co-ordination A, and hence the uniform delay  $d_u$ , is now zero (correct, as no vehicles stop) and with co-ordination being the worst possible or 0% arriving during green, the

uniform delay is double that calculated using W+C (again correct, as double the number of vehicles must stop compared to random arrivals).

Taking it further, it is theoretically possible for  $\mu$  to equal -100% which would represent a situation when the entire platoon arrives exactly at the start of red, and not uniformly throughout the red which 0% represents. This would be equivalent to multiplying the W+C uniform delay by 4, but this is so unlikely that it is not specifically mentioned in the AutoJ User Manual.

#### 6.4.2. RANDOM DELAY

Simulation runs of the W+C formula indicate that the random delay component,  $d_r$  (the second part containing B), is extremely sensitive to the  $g/c$  ratio. Common sense would dictate that while uniform or system delay,  $d_u$ , depends on  $g/c$ , the random delay (which is the delay caused by other vehicles in the traffic stream) should depend on the  $V/C$  ratio rather than the amount of green time allowed. Hence using the W+C formula, a  $g/c$  equal to 0.1 gives a random delay five times higher than with a  $g/c$  of 0.5 even with the  $V/C$  being identical in both cases.

After simulation testing, it was found that replacing the 2 in the divisor of formula B with  $g/c$  in the multiplier cancels out the  $g/c$  sensitivity. If  $g/c$  equals 0.5, then an identical result to the W+C formula occurs, but even if  $g/c$  does not equal 0.5, the random delay component in the modified formula does not change.

It is argued therefore that a more accurate result is achieved if the formula for B is modified to read:

$$B = g/c * x^2 / (1 - x)$$

This can be simplified further by considering  $x = q / (S/3600 * g/c)$  and taking  $S = 1800$  as used by Webster and Cobbe,

$$B = g/c * q / (1800/3600 * g/c) * x / (1 - x) = 2 * q * x / (1 - x)$$

and the second term,  $B / q$  simplifies to

$$d_r = 2 * x / (1 - x)$$

(completely independent of  $g/c$  and  $q$  and only dependent on  $V/C$  as one would expect).

#### 6.4.3. CORRECTION FACTOR

According to W+C, the third term C results in a correction of between 5% and 15%. This term was also extensively simulated. The results indicated that the 0.9 approximation suggested by W+C (which effectively reduces calculated delay by 10%) should only apply to the second part of the formula, part B, as the uniform delay, part A, did not need correction. Therefore the 2 in part B should read 1.8.

#### 6.4.4. FINAL FORMULA

The final formula recommended for delay at traffic signals with a V/C in the range of 0.0 to 0.975 is:

$$d = (1 - \mu) * c * (1 - g/c)^2 / (1 - g/c * x) + 1.8 * x / (1 - x)$$

### 6.5. DELAY AT OVERSATURATED INTERSECTIONS

Neither Webster and Cobbe's (1966) formula, nor the modified formula above, can be used in conditions with a V/C of greater than 0.975. The following alternate formulations were therefore considered for oversaturated conditions:

#### 6.5.1. MODIFIED BUREAU OF PUBLIC WORKS (BPR)

$$T = t_0 * (1 + a * (Q/C_s)^b)$$

where relating this to the symbols defined in 6.2 above, T equals total delay, previously shown as d;  $t_0$  is the initial or uniform delay,  $d_u$ ; Q is the volume V;  $C_s$  is the practical capacity taken at  $0.975 * C$ ; and a and b are constants.

Various values for the constants were tested and it was found that  $a = 7$  and  $b = 4$  gave a reasonable match to the W+C formula in under-saturated conditions.

#### 6.5.2. MODIFIED DAVIDSON

The modified Davidson formulae are:

$t = t_0 * [1 + J * Q / (C - Q)]$	$Q < C_s$
$t = t_0 * [1 + J * (C * Q - C_s^2) / (C - C_s)^2]$	$Q \geq C_s$

where the symbols are defined as for BPR in 6.4.1 above and J is a constant.

After testing, J was set at 0.3.

#### 6.5.3. JHK

The JHK & Associates formula stems from work done on NCHRP 3-82(2) project **Urban Signalized Intersection** (as presented at the Prof Adolf May Course on Highway Capacity, South Africa, July 1982).

$$d_u = 0.385 * c * (1 - g/c)^2 / (1 - V/S)$$

$$d_r = (1500 / 13) * [(x - 1) + ((x - 1)^2 + 12 * (x - OF) / (S * g/c))^{0.5}]$$

$$OF = 0.67 + (c / (3600 * 600)) * S * g/c$$

#### 6.5.4. AUTOJ (SAMPSON 2016)

The modified W+C formula for an  $x = V/C$  of between 0.0 and 0.975 ( $x \leq y$ ) has been derived in section 6.4 above.

For an  $x$  of greater than  $y$  ( $y = 0.975$  is the  $V/C$  at practical capacity), the following formula was derived:

$$d = d_{Cs} + 1800 * (1 - y/x)$$

where  $d_{Cs}$  is the delay at practical capacity and  $1800*(1-y/x)$  is the overflow delay.

This formula was derived by theorizing that in the worst possible case of a capacity of zero, no vehicle can proceed and hence every vehicle arriving at the stop line must be delayed for the full period between arrival time and the end of the hour being modelled.

Consider one hour at a time with no vehicles waiting prior to the beginning of the hour. The first vehicle arriving will wait the full 3 600 seconds. The last vehicle arriving at the end of the hour will wait zero seconds (during that hour). With uniform arrivals, the average delay per vehicle cannot therefore exceed 1 800 seconds, no matter how many vehicles arrive in the hour.

During subsequent hours there will obviously be further delay as the queue at the start of the next one-hour period will equal the number of vehicles that have arrived during the subject hour, but for the moment we are only concerned with the hour with no queue present at the start.

The additional (overflow) delay,  $d_o$ , is therefore limited to an average of 1 800 seconds per vehicle for that first hour but will be less than that if the capacity is not zero.

Using the formula above, at a  $V/C$  of twice the practical capacity ( $x = 1.95$ ), the overflow delay is 900 seconds. To this must be added the delay at practical capacity. This makes sense if it is considered that at a  $V/C$  of 1.95, vehicles arriving during the first half hour will be served during the hour and will have gone by the end of the hour. The average wait for those arriving in the second half hour and not served will be half of 1 800 seconds, or 900 seconds.

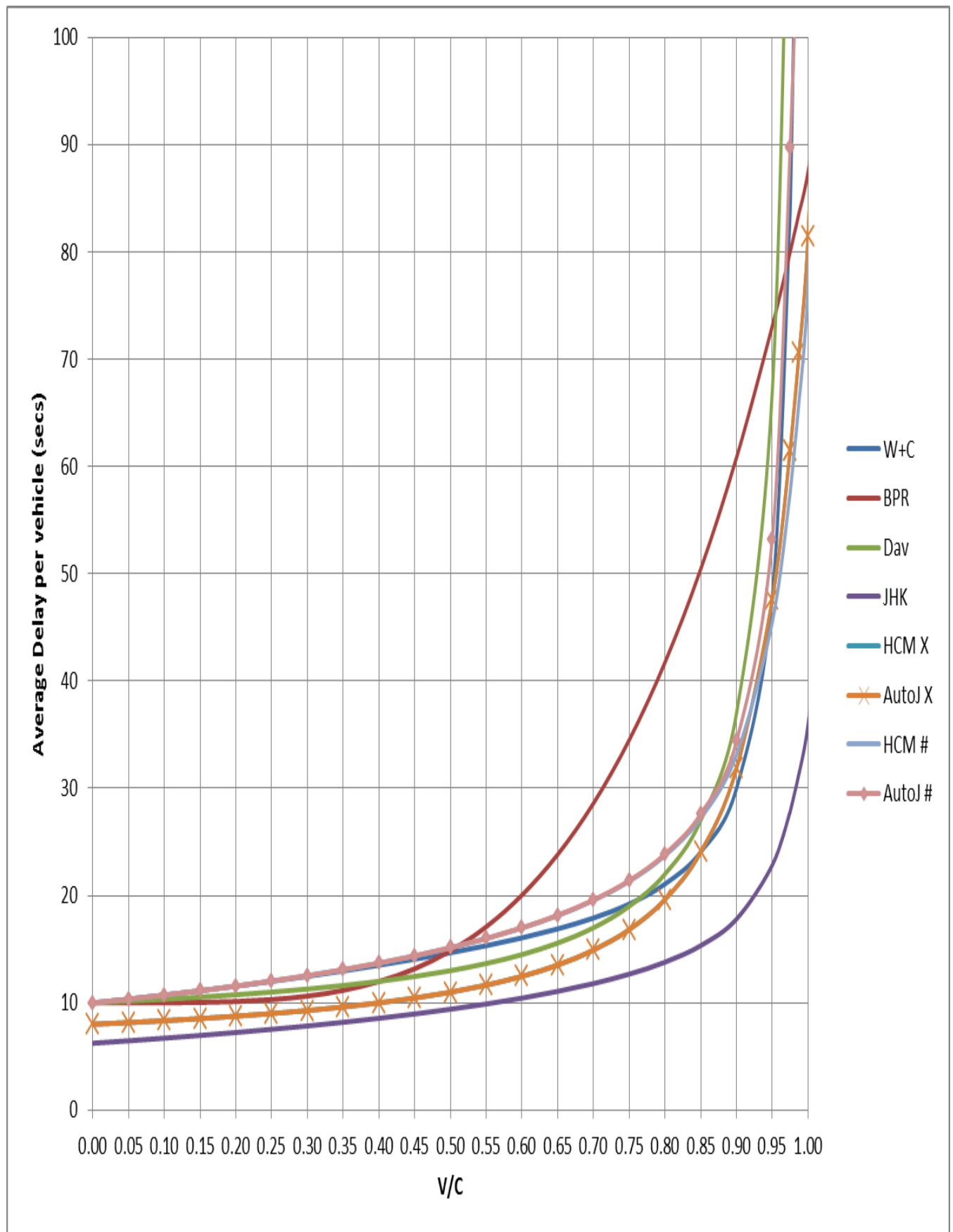
Note that in oversaturated conditions, all the other methods (described in sections 6.4.1, 6.4.2 and 6.4.3) at some point produce average delays per vehicle are well in excess of one vehicle hour per hour which is clearly impossible.

This is a new formulation for overflow delay not seen in other references and quite different from the formulae derived by other researchers above. It is argued however that it is a much more accurate formulation and is recommended for transportation and traffic engineering simulation models.

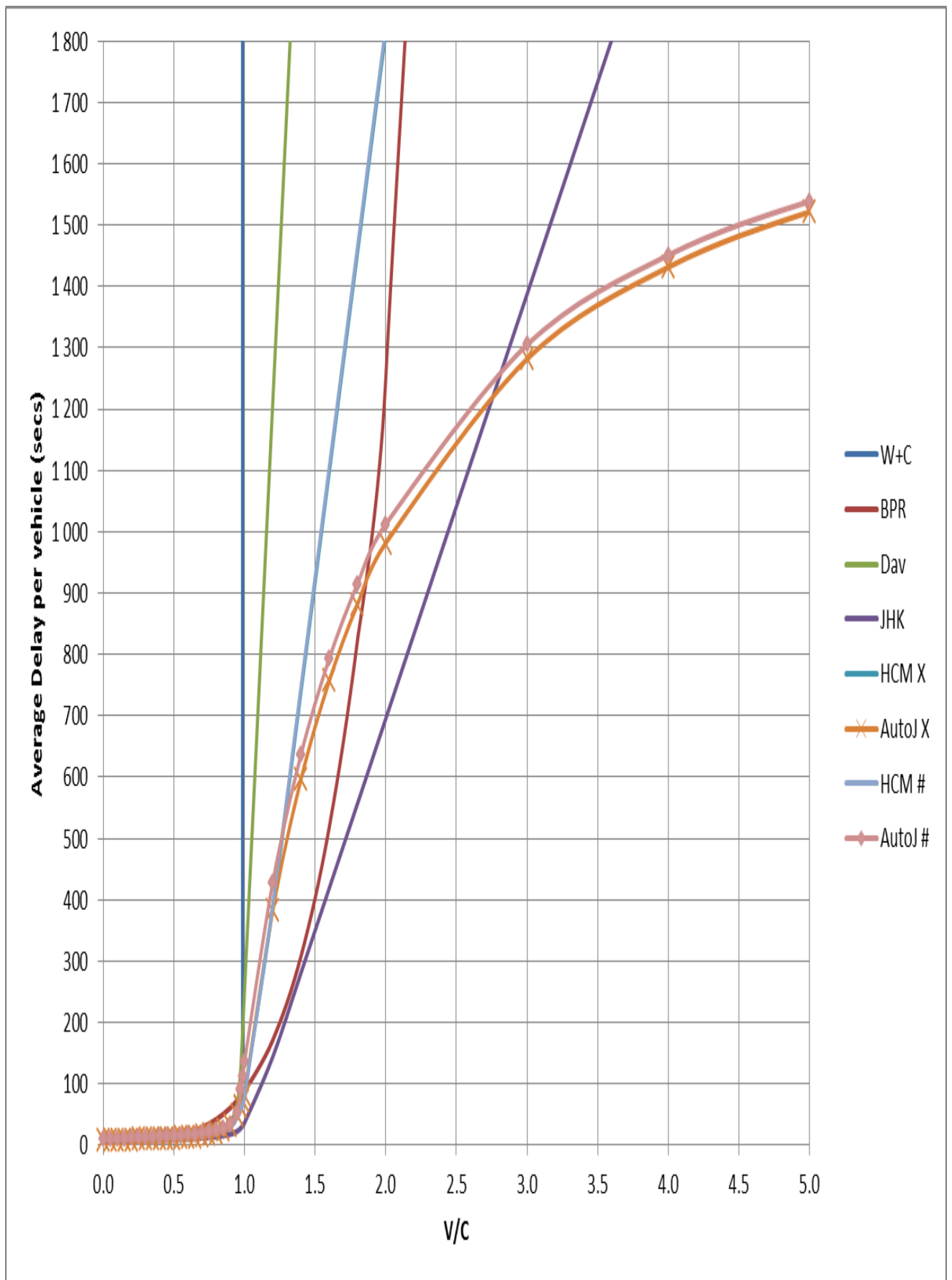


## **6.6. RESULT COMPARISON**

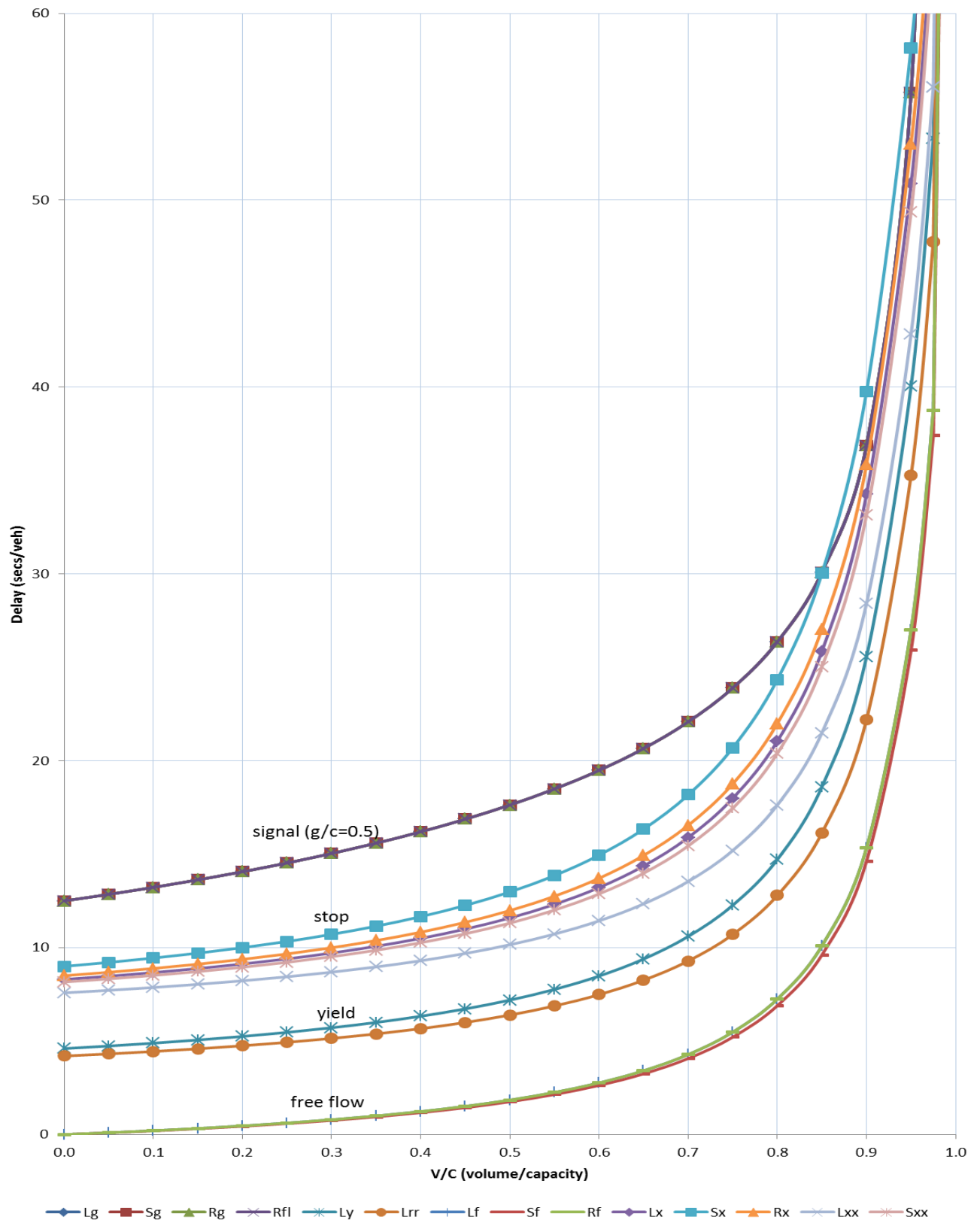
A comparison of the results using the different methods described above with a cycle time of 70 seconds, g/c of 0.5 and a random arrival pattern is given in Figures 12a, 12b, 13a and 13b.



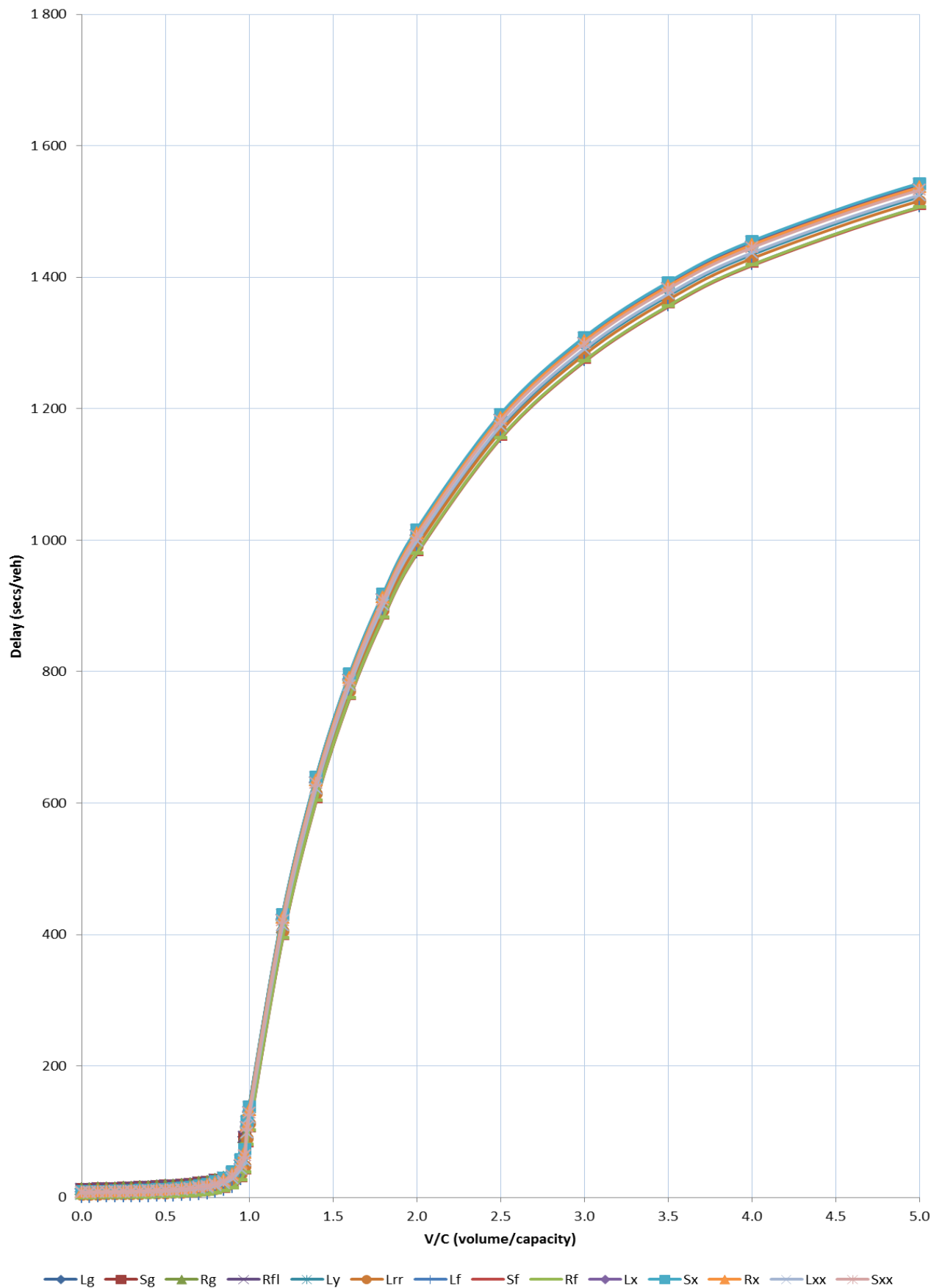
**Figure 12a: Average under saturated delay using different formulae (X=stop, #=signal)**



**Figure 12b: Average delay in over saturated conditions using different formulae**



**Figure 13a: Effect of under saturated volume to capacity ratios on delay with different control devices using the AutoJ formula (Sampson 2016)**



**Figure 13b: Effect of over saturated volume to capacity ratios on delay with different control devices using the AutoJ formula (Sampson 2016)**

## 6.7. CONCLUSION OF DELAY EQUATIONS

For under-saturated conditions, the delays at Stops are calculated using the **HCM** method, while for traffic signals the **Webster and Cobbe** formula has been used as a base. These methods cannot be used in over-saturated conditions however and a new formula has been developed for these situations.

It is argued that the delay equations for both under-saturated and over-saturated demand conditions, derived or amended as above, are more accurate than the formulae found in some references and are not sensitive to non-relevant factors. They also cover the full range of V/C from 0 to infinity without restriction.

## 6.8. QUEUES

The queue length is a function of delay and traffic volume. The average delay per vehicle multiplied by the number of vehicles per hour gives the total delay at the intersection in vehicle-hours per hour. The same multiplication gives the queue length. If the delay slowing down and accelerating is ignored, total delay and total queue are therefore numerically equal.

Thus, if the requirement is to equalize queues on each approach, we cannot simply independently minimize average delay. What is needed is to reduce average vehicle delay to the high-volume movements by increasing the green time with a corresponding decrease in green time to the low volume movements which will increase that delay, until the products are equal. While this could be a legitimate strategy, it is rarely done.

## 7. LEVEL OF SERVICE

### 7.1. LEVEL OF SERVICE RANGE

The **Level of Service** (LoS) is defined by the HCM (TRB 2000 and 2010) and others as follows:

- Priority control, average **delay** (secs / veh): A<10, B<15, C<25, D<35, E<50, F=50+
- Signal control, average **delay** (secs / veh): A<10, B<20, C<35, D<55, E<80, F=80+
- Vehicular **volume / capacity** (V/C): A<0.5, B<0.8, C<0.9, D<0.95, E<0.99, F=0.99+
- Pedestrian **volume / capacity** (V/C): A<0.1, B<0.3, C<0.4, D<0.6, E<0.97, F=0.97+
- Pedestrian **density** (ped/min/m): A<7, B<23, C<33, D<49, E<82, F=82+
- Ped volume (ped/metre/hour) A<400, B<1200, C<2000, D<3000, E<4800, F=4800+

It can be seen therefore that Level of Service is not an independent criterion but is based on other measures. The delay measure is a subjective measure of driver comfort while the V/C measure is a more accurate and scientific measure of actual traffic conditions. While related, they do not always agree.

### 7.2. MEASURES OF EFFECTIVENESS

The performance of an intersection can be judged in different ways. The user therefore must decide which of the following **measures of effectiveness** (MoE), or combination, should be used when optimizing the intersection:

- Volume / Capacity ratio (V/C)
- average Delay
- maximum Delay
- total Delay
- maximum Queue
- total Queue
- Level of Service (based on V/C or delay)

If a movement is operating under heavy load, the volume to capacity ratio becomes important as you would not want any movement to exceed capacity. Under lighter loads, the delay is more important as capacity is unlikely to be of concern. If block lengths are short, the queue may be the most important factor.

The HCM recommends average delay per vehicle as the preferred LoS measure. While in light flow conditions this is best, for reasons given above other LoS measures should also be considered.

Having decided what measure should be used, the next decision is what to optimize:

- a movement
- an approach
- the intersection as a whole

Finally, the period for the optimization needs to be considered:

- quarter hour
- hour (AM, PM or other)
- period (e.g. 3 hours)
- 24 hours

The following observations may assist the decision:

- For an intersection, ideally no movement should exceed a V/C of 1.0 and preferably should be not greater than 0.9, although you may tolerate a V/C of close to 1.0 for a minor movement if major movements benefit;
- An average delay per vehicle of 50 secs at a Stop street is LOS F (TRB 2010) while 57 secs or more is considered "intolerable" in stop street studies (Sampson 1992). Delay will exceed 57 secs in a single lane at a stop street if  $V/C > 0.95$ , hence traffic signals may be specified even if overall delay at the intersection is increased;
- Delays exceeding 50 secs are however common at signals (LOS D), where LOS F is only reached when the average delay per vehicle exceeds 80 secs (TRB 2010);
- The SARTSM 4Q queue length warrant specifies that when a queue of vehicles or pedestrians exceeds an average of four (total delay  $> 4$  veh-hrs / hr) during any hour of the day, a signal is warranted. The average delay that will result in the warrant being met depends on the approach volume. At 100 veh/hr on the side street, average delay / veh = 144 secs, at 200 = 72 secs, at 300 = 48 secs and at 400 = 36 secs etc. Hence delay LoS cannot be used for warranting traffic signals;
- A warrant of 3.0 veh would reduce average delay / veh above to 108, 54, 36 and 27 secs respectively. After extensive experience in implementing the SARTSM warrant, I can confirm that this would be a more realistic warrant based on public and political opinion. However a roundabout is becoming increasingly acceptable and would be a much better alternative to installing signals.



### 7.3. PERFORMANCE INDEX

In AutoJ, the user can weight any of the criteria above, i.e. V/C, delay, queues, movements, time periods or Levels of Service to suit the intersection. The weighting determines the **performance index**.

In general, a higher weighting to delay favours priority control, particularly roundabouts, while a higher weight to V/C favours signals. A higher weighting to movement V/C favours multi-stage signals, while weighting intersection V/C highly favours 2 stage signals.

After much experimentation, the following weightings are recommended to optimize overall control performance throughout the day and night (50% to V/C, 25% to delay, 25% to queue):

- 18% The **V/C** of the worst movement in the AM peak
- 4% The **V/C** of the worst movement in the off peak
- 18% The **V/C** of the worst movement in the PM peak
- 10% The intersection average **V/C** during the worst period
  
- 4% The ave. **delay/veh** in the AM peak
- 7% The ave. **delay/veh** in the off peak
- 4% The ave. **delay/veh** in the PM peak
- 10% The ave. **delay/veh** of the worst movement in any period
  
- 6% The **queue** of the worst movement in the AM peak
- 3% The **queue** of the worst movement in the off peak
- 6% The **queue** of the worst movement in the PM peak
- 10% The total **queue** (total delay) in all periods.

The greatest weight is given to the intersection's V/C performance during peak hours as this is likely to be the most critical measure determine the optimal control.

A higher weighting is given to off peak than peak delay as capacity is less important during that period.

The fact that pedestrians are not weighted does not mean they have not been considered. They feature in the capacity calculations and the signal timings have minimums that always ensure that pedestrians have adequate time to cross. If, however, a weight is put to pedestrian movements as part of the performance index, the effect will be to favour all-way stops because pedestrians have

priority on all approaches and no delay. For reasons stated earlier, this form of ICD cannot be recommended.

#### **7.4. CONCLUSION**

Because the HCM recommends delay as the preferred measure, the other measures are regarded to be of lesser importance, but delay is only one measure of an intersection's performance. It has been demonstrated in the examples in this Chapter that if a combination of measures and the times during which they are important are not considered, the optimal result will not be obtained.

## 8. DESIGN OF TRAFFIC SIGNAL TIMINGS

### 8.1. INTRODUCTION

The formulae for capacity, delay, queues, Level of Service and measures of effectiveness have been examined above. In this section the way in which this knowledge is combined to design a traffic signal staging and timing is described.

### 8.2. CYCLE TIMES

**Cycle times** are generally in the range of 50 seconds to 120 seconds, although cycle times can be shorter or longer in special circumstances. In general cycle times are specified in 10 second intervals but it is possible for controllers to handle 5 second intervals if necessary.

Cycle times are dependent on minimum greens and the number of stages, concepts discussed below. In most circumstances however, cycle times should be kept as short as possible to reduce delay. Longer cycle times may be justified to reduce the percentage of lost time where traffic volumes are high, and the need is to increase capacity, but this benefit is often over-estimated because the small gain in less lost time is cancelled out by longer gaps in the traffic flow.

Cycle times are sometimes set based on system considerations, i.e. for synchronization purposes, all signals in the group must have the same cycle time or an exact multiple thereof.

### 8.3. MINIMUM GREEN, YELLOW AND ALL-RED TIMES

#### 8.3.1. MINIMUM GREENS

In all cases, the minimum green time must at least be equal to the time it takes for a pedestrian to safely cross. This is calculated as the width of the crossing (or the width to a median shelter island) divided by the walking speed of a pedestrian, as per Table 9. Typically, a speed of 1.2m/sec is used.

**Table 9: Typical walking speeds of pedestrians**

Category	(m/sec)	(km/h)
slow (elderly)	1.0	3.6
normal	1.2	4.3
brisk	1.5	5.4

An additional consideration however is that the green time must long enough to give not only the first vehicle but also the next few following vehicles time to enter the intersection. The SA RTSM

recommends a minimum of 11 seconds for the green disc with an absolute minimum of 7 seconds. For flashing arrows, these times can be reduced to 7 seconds and 4 seconds respectively.

### 8.3.2. ACCELERATION AND DECELERATION RATES

Before determining the required yellow and all red intervals, typical acceleration and deceleration rates need to be established. Table 10 gives some typical values as well as the distances and times required to accelerate to, or stop from, 60 km/h on a flat 0% grade.

**Table 10: Acceleration / deceleration rates with corresponding distances**

60km/h and 0% grade with reaction time from start of yellow signal	accel (+), decel (-) rate (m/s <sup>2</sup> )	react time (secs)	dist. travel while reacting (m)	time to / from 60 km/h (sec)	distance to reach 60 km/h or stop (m)	warn time to stop at stop line	total dist needed to stop (m)
<b>ACCELERATION</b>							
Formula 1 race car	14.2	n/a		1.2	9.8		
Typical car	2.0	n/a		8.3	69.4		
Typical truck	0.5	n/a		33.3	277.8		
<b>DECELERATION</b>							
Slowing in gear	-0.7	0.75	12.5	23.8	198.4	12.7	210.9
Comfortable braking	-1.7	0.75	12.5	9.8	81.7	5.7	94.2
Limit comfortable braking	-2.5	0.75	12.5	6.7	55.6	4.1	68.1
Truck stop on yellow signal	-3.1	0.75	12.5	5.5	45.5	3.5	58.0
Car stop on yellow signal (SARTSM)	-3.7	0.75	12.5	4.5	37.5	3.0	50.0
Legal requirement for cars (SA)	-5.8	0.75	12.5	2.9	23.9	2.2	36.4
Cars, expert or with ABS	-9.0	0.75	12.5	1.9	15.4	1.7	27.9
Formula 1 race car, or crash into wall	-50.0	0.75	12.5	0.3	2.8	0.9	15.3

### 8.3.3. MINIMUM YELLOW AND ALL-RED

The **yellow** and **all-red** times are set using the speed, grade and clearance distances from the RTSM formulae, with a minimum yellow interval of 3.0 seconds and a minimum all-red interval of 2.0 seconds.

As can be seen in Table 10, a car can stop at a yellow signal if at least 50m back from the stop line when the yellow appears. This is the same distance it will cover to reach the stop line at 60 km/h during the 3 sec yellow interval if it does not decelerate. This is not a coincidence. It is designed so that whether the car stops or goes, both can be done legally. If the car is further back than 50m when the signal turns yellow, it must stop or else it will enter the intersection on red. If it is closer than 50m when the yellow appears, it must not stop because if it tries it will not stop before the stop line.

In a standard situation (flat grade, 60 km/h), if the road authority decides to extend the yellow beyond 3.0 seconds, a situation occurs where a car can choose to stop or go. In general, as motorists learn of this, they will choose to go rather than stop. Hence, they will violate the legal requirement to stop on a yellow light if they can safely do so. For this reason, the yellow interval is kept as short as possible.

Consider however a truck. If the truck is between 50m and 58m from the stop line (Table 10), it cannot stop before the line, but if it goes it will enter the intersection after it has turned red. This is the **dilemma zone**.

To deal with the dilemma zone, an all red period is required. This allows a vehicle that cannot stop on yellow to safely proceed through and clear the intersection before the other side is given the green light. For wide intersections the all-red is extended beyond the 2 sec minimum.

The dilemma zone would also apply to light vehicles approaching at faster than 60km/h or approaching on a down grade. Therefore, in addition to all-red period, SARTSM allows for longer yellow times in those circumstances.

Furthermore, because there are situations where a vehicle, e.g. a truck, cannot stop during the yellow interval and will legally cross the stop line after the signal has turned red, **enforcement** should not start until the last second of the all red interval.

The all-red interval is also used by right turning vehicles waiting in the intersection at the end of the green phase, but this movement does not influence the duration of the all-red period.

Note that for leading greens the all-red clearance period can be one second less than the main phase all-red in accordance with the RTSM. This is because motorists turning right do not need to completely clear the intersection before the opposing green is displayed.

For safety and efficiency reasons, there should be no all-red between a main stage and a lagging green flash in the same direction. This is to avoid the confusing time gap between these two phases. If there is a gap, motorists waiting for the flash are not sure whether it will be displayed. The uncertainty and confusion can lead to collisions. Having no all-red in this situation is not

referred to in the RTSM which means some authorities will default to a minimum of two seconds. Users should use their own engineering judgement when determining what is best.

#### **8.4. NUMBER OF STAGES**

At the beginning of this report, the standard signal staging options were described. A count of these will show there are 21 sensible traffic signal options. There are other options, such as double leading right turns, or five or six stage options, but for reasons explained elsewhere in this report these are less optimal.

The choice of which of these options to adopt will depend on the traffic volumes, the capacity of each movement, and the Level of Service that will result. To determine this, the calculations of these factors must be made.

There are however some simple considerations that can help in the decision.

The first is that each additional stage will result in additional **lost time**. This lost time is the inter-green time of 3 seconds of yellow plus 2 seconds of all-red, i.e. 5 seconds, or longer if the yellows and all-reds are extended. Hence, a two-stage signal has at least 10 seconds of lost time, a three-stage signal 15 seconds and a four-stage signal 20 seconds of time lost each cycle.

If more than four stages are used, there will be even more lost time. Furthermore, if some stages are repeated in an unexpected sequence, hesitation and uncertainty will result. Therefore, in practice, providing more than 4 stages is inefficient, confusing and counter-productive.

Another consideration is that if the cycle time is reduced, the need for additional stages can be avoided. For example, with a 60 second cycle, there are 60 cycles and therefore 60 inter-green periods per hour. With a 120 second cycle, there are only 30. Typically, 2 vehicles can turn right at the end of each stage, hence with a 60 second cycle, 120 vehicles can turn on the inter-green; but with a 120 second cycle this is reduced to 60. Hence if between 60 and 120 vehicles per hour turn right, they can comfortably do this during the inter-green if the cycle is 60 secs, but will require an additional stage if the longer cycle is adopted.

#### **8.5. SIGNAL TIMING DESIGN**

The required steps to obtain accurate signal settings are described in this section.

##### **8.5.1. OVERFLOW VEHICLES**

When auxiliary left and right turning lanes are too short, one of two things can happen; 1) turning vehicles are blocked from entering the lane by the straight vehicle queue, or 2) the turning queue will spill back into the adjacent lane.

To avoid the problem of overflow, the length of the **required** turning lane must not be less than the 85<sup>th</sup> percentile queue. Based on the Highway Capacity Manual (TRB 2000) recommendations for priority control, this is taken as the volume arriving in the turning lane every two minutes multiplied by an average vehicle length (a passenger car length plus gap is typically 6.0m).

In AutoJ, this is then compared to the **actual** auxiliary lane length available (defined as widened length plus half the length of the taper). If the **actual** auxiliary lane (left or right turning lane) length is more than the **required** length no adjustment is necessary. If the actual length is shorter, the effective volume using the adjacent lane is higher than the actual volume.

The formula used in AutoJ (Sampson 2016) for determining the average **capacity** of a turning lane is the number of vehicles that can fit into the lane per hour if they fully discharge every minute:

$$C_o = \text{actual lane length} * \text{no. of lanes} / \text{vehicle length (default 6.0)} * 60 \text{ veh/hr.}$$

The overflow volume, if any, is the actual turning volume less  $C_o$ . This overflow volume must be added to the adjacent straight lane volume to get an equivalent or effective volume for timing calculation purposes, assuming the traffic signal does not have a flash for that lane. In effect, if the capacity of the turning lane is inadequate, more green time will be required on that approach to try and clear the overflow.

The reason why a one-minute discharge is used for capacity in the formula above, and not two minutes as for the HCM recommendation, is because the 85<sup>th</sup> percentile length expected with random arrivals is approximately double the average lane usage.

#### 8.5.2. EFFECTIVE NUMBER OF LANES AND LANE BALANCE

In a multi-lane situation, vehicles do not distribute themselves perfectly equally in each lane, even when travelling straight through. Furthermore, multiple turning lanes suffer from the effect of an adjacent turner blocking visibility and freedom of movement of the turner alongside.

The following formula, based on the HCM (TRB 2000) method, is therefore applied for traffic distributing itself in lanes in multi-lane situations at traffic signals:

$$N_{eff} = (N_{actual})^x$$

where  $N$  = number of lanes and  $x = (1+0.885)/2 = 0.94$ ,  $(1+0.952)/2 = 0.98$  and  $(1+0.971)/2 = 0.99$  for left, straight and right turns respectively.

For roundabouts, stops, yields and all-way stops, there are no HCM figures. In earlier versions of AutoJ, the values for  $x$  for these ICD's were set by the author at 0.92, 0.70, 0.70 and 0.92 respectively. The reason these were set at lower than the HCM values is that at priority controls it

can be observed that when visibility in a lane is obscured by adjacent vehicles, those lanes are avoided.

The effects of these adjustments are shown in Table 11, but this is for reference only as later versions of AutoJ calculate capacity on a lane by lane basis and do not use adjustment factors.

**Table 11: Effective number of lanes**

	Lrr	Ly	stop	L only	LS	S only	L(S)R	SR	R only
<i>Power x</i>	<i>0.92</i>	<i>0.92</i>	<i>0.70</i>	<i>0.94</i>	<i>0.92</i>	<i>0.98</i>	<i>0.90</i>	<i>0.90</i>	<i>0.95</i>
<i>penalty</i>					4.0%		8.0%	10.0%	
Actual lanes	Effective number of lanes								
	L r/about	L slip	stop	L only	LS	S only	L(S)R	SR	R only
1	1.00	1.00	1.00	1.00	0.96	1.00	0.93	0.91	1.00
2	1.89	1.89	1.62	1.92	1.82	1.97	1.73	1.70	1.93
3	2.75	2.75	2.16	2.82	2.64	2.92	2.49	2.44	2.84
4	3.58	3.58	2.64	3.69	3.44	3.87	3.22	3.17	3.73

### 8.5.3. MIXED LANES

Mixed lanes (combinations of left, right and straight traffic in the same lane) add another factor of imbalance. Straight through vehicles tend to avoid the lane used by turning vehicles even if it is legal to use it.

Where lanes are shared therefore, an adjustment needs to be made. In a shared right and straight lane for example, it is highly probable that straight through vehicles will be held up by vehicles waiting to turn right in the same lane. Where there is not a flash therefore, in AutoJ the capacity of the shared lane was determined by the demand volume with 90% of the volume contribution allocated to the lower capacity movement and 10% to the higher capacity.

### 8.5.4. OPPOSING VOLUME ADJUSTMENT

The way in which equivalent opposing volumes are calculated was shown in 4.2.1. *Derivation of  $V_c$  the conflicting flow*. At a traffic signal however, the opposing volume is concentrated as it can only proceed when the signal is green. To determine the availability of gaps therefore, the



opposing volume must be divided by the green to cycle (g/c) ratio, which results in a much higher effective opposing volume.

However, this gives rise to a difficulty. To determine the g/c ratio, the volume to capacity (V/C) of the critical movement is needed, but the capacity is in large part determined by the number of vehicles that can take gaps. So, the number of gaps cannot be determined until the g/c is known and vice versa.

To overcome this difficulty, the planning method is used to determine the initial g/c. This method assumes that a right turn vehicle is equivalent to two straight through vehicles.

## 8.6. CRITICAL LANES

Traffic signal **splits** are generally determined using the **Critical Lane Method**. This simply means that when determining the green time, select the lane with the highest **critical volume** which utilises each green phase, and use that to calculate the green split.

Traditionally, **critical volume** was taken to be the actual volume multiplied by the difficulty rating. In the “planning” method, a right turn is assumed to be twice as difficult as the straight or left, so the critical lane will be the highest of the left, straight or two times the right volume per lane. This crude but simple method can be used for the initial setting of signal green times.

The more accurate and correct method is described in the next section.

### 8.6.1. VOLUME PER PHASE

Straight through vehicles can only proceed when the green disc is displayed. Left turn vehicles similarly, unless there is a left turn flash when they can also utilize that green time. Right turn vehicles have three options, 1) taking gaps during the green disc phase, 2) turning on the inter-green period, and 3) turning on the flash if one is provided.

For signal setting purposes, the right turn volumes utilizing each of these options is allocated as follows:

- The number of vehicles that can turn during inter-green is calculated. This is then multiplied by the V/C ratio on the assumption that with spare capacity, most right turners will take gaps whereas as capacity is reached, the inter-green capacity will be fully utilized.
- This number is subtracted from the number of vehicles turning right. The reason for doing this is that these vehicles will be able to turn right regardless of the signal green time. If they are not subtracted, AutoJ will assume they are still waiting would allocate more and

more green time for that movement. The problem of allocating excessively long green times for minor movements is therefore avoided.

- If there is no flash, the volume utilizing the green disc phase is therefore the demand volume less those turning on inter-green.
- If there is a flash, the volume using the flash is taken as the demand volume less the inter-green turners less 95% of those vehicles that can take a gap. The 95% is assuming some vehicles that might have taken a gap will prefer to wait for the flash.

Note that the above procedure was used only to determine signal timing. When calculating the actual V/C the full right turning volume is divided by the inter-green capacity plus the taking gaps capacity multiplied by the green disc g/c plus the flash capacity multiplied by the flash g/c.

#### 8.6.2. VOLUME PER LANE

In earlier versions of AutoJ the volume in each lane was calculated. A one step iterative process was followed whereby all movements were initially allocated equal volumes in each available lane. The V/C of each lane is then calculated. As some lanes will have a higher V/C than others, some of the movement volume was reallocated to the lane with the lowest V/C until the V/C in each lane was approximately equal. The reallocated volume per phase and per lane was then used to calculate the critical volume.

This method was abandoned however when it was found to be more accurate to calculate the total capacity available to each movement, including the share of mixed lane capacity, and to take the total movement demand volume and divide by that capacity. This calculation also means that each movement can have different levels of service, as happens in practice, whereas the previous calculation method would allocate the same level of service to all movements sharing a lane.

#### 8.6.3. OPTIMIZATION METHOD

The downside of the critical lane method is that a minor flow with a very low capacity (and hence high V/C) may be given a disproportionate amount of green time. In practice this means that while the V/C of each critical movement is approximately equalized by sharing out the green time in that proportion, it can result in reduced green time for higher volume movements with additional delay and a higher V/C for the intersection as a whole.

To overcome this, a method to favour higher volume critical lanes was experimented with (Sampson 2016). An optional formula for critical lanes is:

$$CL = V / (C / S)^x; \text{ where:}$$

CL = critical lane volume

V = actual volume

C = “capacity” with 100% green time, calculated as described above

S = saturation flow, 2000 per lane by default

x = power factor.

The formula was developed using simulation. Its effect using  $x = 0.67$  is contrasted with the traditional method in Table 12 below (relative, not absolute values are listed).

**Table 12: Relative V/C versus C/S multipliers for critical lane volumes**

V/C (C/S) <sup>1.0</sup>	1.0	1.5	2.0	3.0	4.0	5.0	10.0
(C/S) <sup>0.67</sup>	1.0	1.3	1.6	2.1	2.5	2.9	4.7

If the V/C is 10.0, the critical volume using the V/C multiplier is 10 times the original volume. Using the amended method with  $x = 0.67$ , this is reduced to 4.7 times.

A power factor of 0.67 was used in earlier AutoJ versions; but this approach was abandoned to avoid the problem of unequal V/C ratios on different approaches.

The critical lane formula now used is:

$$CL = V / (C / S) = V / C * S.$$

## 8.7. WEIGHTING FAVOURED MOVEMENTS

In AutoJ (Sampson 2016) a weighting factor is provided to allow the user to force the program to give more (or less) green time to chosen movements, for example to favour a bus lane. The weighting factor is multiplied by the actual volume to give a higher (or lower) effective volume for the weighted movement. This however will only work if the favoured lane was a critical lane, or after weighting it becomes a critical lane.

Alternatively, the user can adjust the green times manually to favour the preferred movements. In both cases the V/C ratio and the delay to the non-favoured critical movements will increase if this is done.

## 8.8. PEDESTRIAN SIGNALS

### REQUIREMENT FOR PEDESTRIAN HEADS

Pedestrian signals (green man, flashing red man, steady red man) reduce the time available to pedestrians to enter a crossing. When crossing on a green disc, pedestrians get the same green time as vehicles. When pedestrian heads are installed they are restricted to only enter the crossing during the green man time.

Pedestrian heads therefore are not normally needed at a two-stage signal unless the crossing distance is so long that the vehicle yellow and all red provides insufficient warning time for pedestrians to clear safely.

Also pedestrians are notorious for ignoring the pedestrian signals so installing pedestrian heads may be fruitless in situations where they will be disregarded.

For the above reasons, pedestrian heads for 2 stage signals need not be provided unless crossing with the vehicle displays is unsafe.

When a flashing right or left turn green arrow exists however, if pedestrians assume they will get the next green they may start to cross without realizing motorists have right of way. Hence pedestrian heads should always be provided on all the approaches affected when the flash gives motorists right of way, particularly if the flash is leading.

An alternative to pedestrian heads is to ban pedestrians from crossing where the flash conflicts by using “no pedestrians” R218 signs and not paint pedestrian crossing lines.

#### 8.8.1. RESTRICTED GREEN FOR PEDESTRIANS

Pedestrians should be given as much green man as possible. This means the green man should be the total time available less the pedestrian crossing time. The practice of providing a deliberately short green man (followed by the minimum flashing red man and then a steady red man while the green disc for vehicles is still displayed) is not encouraged. This attempt to give turning vehicles more time to turn a) does not work very well, b) will probably be ignored, and c) the steady red man is meant to prevent pedestrians crossing when it is illegal and unsafe, not to allow vehicles priority over pedestrians.

#### 8.8.2. EARLY START FOR PEDESTRIANS

Another practice with doubtful benefits is to start the pedestrian green man between one and three seconds before the green disc for motorists. Pedestrians in any event have priority and giving them an early start is not necessary. Furthermore, if the practice of giving pedestrians an early start becomes widespread, it could almost be taken as encouragement to motorists to not give priority to pedestrians where the early pedestrian start is not provided.

### 8.9. CO-ORDINATED FIXED TIME OR VEHICLE ACTUATION

For a variety of operational reasons, the rule for deciding whether fixed time or vehicle actuated signals are better is simple:

- if a traffic signal is isolated (further than one kilometre from the next nearest signalized intersection) or if it is not on a Class 1, 2 or 3 arterial mobility road, it should be fully vehicle and pedestrian **actuated**;
- in all other cases, it must be **co-ordinated / synchronized** with a fixed cycle time.

On occasions where unwarranted signals exist on arterials (an unfortunate situation often found when a developer pays for a signal) or when the signal is only needed during part of the day (e.g. near a conference centre), then the side road signal must be **semi-actuated** in addition to being co-ordinated. The arterial stays green unless there is a demand from the side street; on demand the side street must wait for the co-ordination pulse so that the side street, and not the arterial, is delayed.

## MAIN REFERENCES

The following major references quoted in this report. References to other works not specifically quoted are attached after the Appendices.

Google [2016] *Lagging versus Leading Green Safety*: Collection of References found on Google, August 2016 (see Annexure)

Joubert, HS, [2010] *Capacity Analysis of Priority Controlled Intersections including Roundabouts*, (notes regarding HCM 2000 to Kittelson and Associates and SC van As), Pretoria, 2010

Rodegerdts, LA, [2015] *Reassessment of Roundabout Capacity Models for the HCM*, Slide presentation downloaded, Kittelson and Associates, 1 May 2015

Sampson, JD, [1983] *Traffic Flow Variations in Urban Areas*, Proceedings of the Annual Transportation Convention, August 1983

Sampson, JD, [1992] *Warrants for Traffic Signals*, PhD Thesis, University of the Witwatersrand, Johannesburg, 1992

Sampson, JD, [2016] *AutoJ 1608, Technical Manual*, and repeated in *SARF Optimizing Intersections Course Notes*, Sandton, South Africa, August 2016.

Sampson, JD, [2017] *Traffic Flow Variations in Urban Areas*, Southern African Transportation Convention, July 2017

Transportation Research Board, [1975] *Traffic Flow Theory*, Special Report 165, Washington DC, 1975, as updated

Transportation Research Board, [1994] *Highway Capacity Manual*, Special Report 209, Washington DC, 1985 (1994 update)

Transportation Research Board, [2000] *Highway Capacity Manual*, Washington DC, 2000.

Transportation Research Board, [2010] *Highway Capacity Manual*, Washington DC, 2010.

van As, SC and HS Joubert, [1993] *Traffic Flow Theory*, University of Pretoria, 1993.

Webster, FV and Cobbe, BM, [1966] *Traffic Signals*, Road Research Technical Paper No. 56, Road Research Laboratory, Ministry of Transport, UK, 1966.