

**Adaptive Urban Bimodal (Car and Transit)
Transportation System Management
in Saturated and Oversaturated Conditions**

A thesis submitted to
the Faculty of Graduate Studies and Research
in Partial Fulfillment of the requirements for the degree

Doctor of Philosophy

by

Kornél Mucsi

BSc, University of Novi Sad, 1984
MSc, Polytechnic of Central London, 1990
MAsc, University of Toronto, 1994

Department of Civil and Environmental Engineering
Carleton University
February 2009

The Doctor of Philosophy Program in Civil Engineering is a joint program with the
University of Ottawa, administered by the Ottawa-Carleton Institute of Civil Engineering

© 2009 Kornél Mucsi



Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-52071-0
Our file *Notre référence*
ISBN: 978-0-494-52071-0

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Abstract

This research focuses on traffic and transit operation (bimodal) in mixed flow congested corridors. To increase transit ridership and modal share, the relative advantage of transit travel times compared to car travel times has to improve. To achieve this objective, this research focuses on three areas of traffic management, the philosophical framework, the development of traffic engineering control techniques, and the evaluation framework. These three areas are tightly interconnected and the successful implementation of bimodal transportation management is not likely without a coordinated development of all three.

The prevailing philosophical framework that provides the foundation to most traditional transportation planning and traffic engineering theories and practices has been reviewed. The research concludes that a number of aspects of the philosophical framework will have to change to address the problem of traffic management in congested conditions in an effective way.

The main focus of this research is on the development of bimodal traffic engineering control techniques for saturated and oversaturated conditions that improve the travel time advantage of transit vehicles while maintaining the existing vehicular capacity of the roadway infrastructure. These measures belong to the family of transit queue jumps but they are more complex than the typical queue jump because they involve queue management as well. A number of signal control algorithms are developed for the

purpose of queue management. An innovative approach, based on fuzzy logic, is also used in some of the control algorithms.

Queue management is a dynamic process and requires continuous data collection. Data collection is not an issue in a simulation environment, but gathering reliable data for real-world applications is. A method based on fuzzy logic and ANFIS (Adaptive Neuro Fuzzy Inference System) tuning technique using detector occupancy data is developed for assessing the number of vehicles approaching an intersection.

The evaluation and justification of bimodal traffic management measures is problematic with traditional traffic engineering or transportation planning performance measures.

Traditional traffic engineering performance measures (e.g. mean vehicle delay) have a single mode perspective, and existing transit or transportation planning related measures of performance (e.g. transit modal share) are not directly linked to traffic operations.

Performance measures designed for the evaluation of bimodal traffic control techniques in saturated and oversaturated conditions are proposed.

Acknowledgements

I am deeply indebted to my supervisor Prof. Dr. Ata Khan whose guidance and encouragement helped me during the entire process. I would like to thank my employers, the City of Ottawa and iTRANS, for their support.

Special thanks go to my family, particularly to my wife Ági, who supported me during all these years with her unconditional love and for keeping our family functional. My children Gergely, Lórinç, Zoltán, and Julianna deserve more ice cream than I will ever buy to them. Finally, I would like to thank to my in-laws Dönci and Kati for taking over a significant part of my role in the family during the last eight years.

Table of Contents

1. Introduction	1
1.1 Overview.....	1
1.2 Problem Statement.....	5
1.3 Research Objective and Scope.....	6
1.3.1 Why is this Research Different from Others? What is the Contribution to Knowledge?	8
1.3.2 Limitations of the Proposed Approach	11
1.4 Study Methodology	11
1.5 Organization of the Thesis.....	15
2. Background.....	18
2.1 Introduction.....	18
2.2 Definitions	18
2.2.1 Operational Consequences of Congestion	20
2.3 Traffic Signal Control: A Brief History.....	26
2.4 Traffic Signal Operation	28
2.4.1 Signal Timing/Optimization Models	32
2.4.2 Traffic Signal Control Systems.....	37
2.5 Traffic and Signal Operation Simulation Tools.....	41
2.6 Summary and Conclusions	41
3. Literature Review	43

3.1	Introduction.....	43
3.2	Approaches to deal with Congested Conditions.....	43
3.2.1	Traffic and Signal Operation Analysis Models for Congested Conditions.....	43
3.2.2	Approaches to Traffic Control/Management in Saturated and Oversaturated Conditions	46
3.2.3	Pre-Signals and the Gating Logic	57
3.3	Fuzzy Logic in Traffic Control.....	63
3.3.1	Philosophical and Historical Aspects of Fuzzy Logic	63
3.3.2	Fuzzy Logic in Traffic Control	66
3.4	Metering.....	73
3.5	Summary and Conclusions	78
4.	The Philosophical Framework of Bimodal Transportation Management.....	80
4.1	Introduction.....	80
4.2	The Need for a New Philosophical Framework	80
4.2.1	Traffic Management vs. Traffic Accommodation	80
4.2.2	The Connection between Long-Term Transportation Planning and Short-Term Operations Objectives	89
4.2.3	Improvements to All vs. Relative Transit Improvements	90
4.3	The Foundations: Why does Bimodal Queue Management Work?.....	91
4.3.2	Discussion and Conclusions.....	98

5.	Bimodal Performance Analysis Framework.....	100
5.1	Introduction.....	100
5.2	The Connection between Objectives and MoP.....	100
5.2.1	“Gaining Efficiency by Abusing Efficiency Paradox” and the “Propagation of Inefficiencies Trap”.....	105
5.3	The Applicability of Traditional MoP in Bimodal Congested Conditions...	116
5.3.1	Mean Vehicle Delay.....	116
5.3.2	Transit Modal Share (TMS).....	138
5.4	Proposed Performance Measures.....	155
5.4.1	Single Mode Perspective.....	156
5.4.2	Bimodal Transportation System Perspective.....	158
5.4.3	Transit System Operations Perspective.....	159
5.4.4	Transportation Policy/Planning Perspective.....	160
5.5	Summary.....	162
6.	Development of Fuzzy Logic Based Queue Management.....	164
6.1	Introduction.....	164
6.2	Motivation to Use Fuzzy Logic.....	164
6.3	The Elements of Fuzzy Logic Control.....	166
6.3.1	Fuzzification.....	169
6.3.2	Rule Base.....	175
6.3.3	Inference and Defuzzification.....	179

6.4	Development of Queue Management Based on Fuzzy Logic	186
6.4.1	Definition of Input and Output Variables	187
6.4.2	Fuzzification.....	188
6.4.3	Definition of the Rule-Base	193
6.4.4	Inference and Defuzzification	196
6.5	Tuning of Fuzzy Systems	204
6.5.1	Tuning Objectives	204
6.5.2	Tuning Techniques.....	206
6.6	Discussion and Conclusions	215
7.	Development of Bimodal Queue Management Concepts	216
7.1	Introduction.....	216
7.2	How to Control the Formation of Queues	217
7.2.1	Queue Jump and Queue Management at the Bottleneck Intersection.....	217
7.2.2	Queue Jump Upstream from the Bottleneck Intersection	221
7.3	Identification of Queue Management Locations	223
7.3.1	Operations Analysis in Unsaturated and Saturated Conditions	223
7.3.2	Operations Analysis in Oversaturated Conditions	226
7.4	Summary and Conclusions	237
8.	Bimodal Queue Management with Partial Queue Jumps	239
8.1	Introduction.....	239
8.2	Description of Roadway Configurations	239

8.2.1	Physical Set-up and Operational Characteristics before Queue Management	239
8.2.2	Physical Set-up for Queue Management.....	240
8.2.3	Relevant Design Input Parameters	244
8.3	Queue Management with Partial Queue Jump and no Pre-Signal (QMnoPS)	246
8.4	Partial Queue Jump with “Park View” Type Control.....	247
8.4.1	Algorithm Description and Analysis.....	247
8.4.2	Algorithm Tuning	253
8.4.3	Control Performance of the “Park View” Type Algorithm	256
8.5	Partial Queue Jump with “On-Off” Type Control.....	257
8.5.1	Algorithm Description	257
8.5.2	Control Performance of the “On-Off” Type Algorithm.....	258
8.6	Partial Queue Jump with Model Based Control	259
8.6.1	Model description.....	259
8.6.2	Control Performance of the Model Based Algorithm.....	261
8.7	Partial Queue Jump with Fuzzy Logic based Control	263
8.7.1	Algorithm Description	263
8.7.2	Tuning the Fuzzy Inference System.....	269
8.7.3	Control Performance of the Fuzzy Logic Based Algorithm	272
8.8	Comparison of Control Algorithms for Queue Management with Pre-Signal	273
8.8.1	Geometric Set-up and Input Data.....	274

8.8.2	Evaluation Results.....	276
8.8.3	Discussion	280
8.9	Simulated Real-World Example: Carling Bronson	283
8.9.1	Corridor Description and Operations for the “Before” Scenario	283
8.9.2	Geometric Set-Up and Operation of Partial Queue Jump without Pre-Signal: Scenario A	288
8.9.3	Geometric Set-Up and Operation of Partial Queue Jump with Pre-Signal: Scenario B	291
8.9.4	Simulation Results	294
8.9.5	Single Mode Perspective.....	295
8.9.6	Bimodal Transportation System Effectiveness Perspective: Impact on Person Travel Times	299
8.9.7	Transit System Operations Perspective.....	300
8.9.8	Transportation Policy Perspective.....	303
8.10	Discussion and Conclusions	307
9.	Queue Relocation to the Upstream Main line Approach.....	308
9.1	Introduction.....	308
9.2	Description of Roadway Configuration.....	308
9.2.1	Physical Set-Up and Operational Characteristics before Queue Management	308
9.2.2	Physical Set-Up for Queue Management.....	309

9.3	Development of Queue Management Algorithm	310
9.3.1	Tuning and Simulation	320
9.4	Simulation of a Real-World Example: Woodroffe Avenue Corridor.....	324
9.4.1	Existing Corridor and Operations	324
9.4.2	Proposed Geometric Set-Up and Operation.....	338
9.4.3	Simulation Results	341
9.5	Summary.....	361
10.	Estimation of the Number of Vehicles in the Detection Zone	362
10.1	Introduction.....	362
10.2	The Need for Estimation of NVDZ	362
10.3	Motivation for Fuzzy Logic Based NVDZ Estimation.....	366
10.4	Fuzzy Inference System Development for NVDZ Estimation	374
10.4.1	Network Based Learning Systems.....	374
10.4.2	Adaptive Neuro-Fuzzy Inference System (ANFIS)	385
10.4.3	The Application of ANFIS for Parameter Tuning	386
10.5	Evaluation of Fuzzy Logic Based NVDZ Estimation	391
10.5.1	Evaluation Results	393
10.6	Discussion on Using ANFIS for Determining the NVDZ.....	418
10.7	Simulation of Real-World Example: Queue Management with Fuzzy Logic Based Estimation of NVDZ	419
10.7.1	Geometric Set-Up and Operations Description	419

10.7.2	Simulation Process.....	420
10.7.3	Evaluation of Simulation Results	423
10.7.4	Analysis of the Control Algorithm Performance.....	425
10.8	Summary.....	427
11.	Final Discussion, Conclusion, and Recommendation	428
11.1	Discussion and Conclusions	428
11.1.1	Changing the Philosophical Environment of Transportation Planning and Traffic Control	429
11.1.2	Development of Bimodal Traffic Management Techniques	432
11.1.3	Bimodal Performance Measurement	435
11.2	Recommendations for Future Research.....	437
12.	References	440
13.	Appendixes.....	462

List of Tables

Table 4.1:	The “Commuter’s Dilemma”.....	90
Table 4.2:	Conceptual Analysis of MoP.....	97
Table 5.1:	Green Time and Delay Calculation for Example 1	109
Table 5.2:	Green Time and Delay Calculation for Example 2	112
Table 5.3:	Suggested Analysis Periods by the Canadian Capacity Guide (1995)	128
Table 5.4:	Input and Output Values for the Deterministic Queuing Model	130

Table 5.5: Mean Delay Calculated by the Deterministic Queuing Model and Analytical Formulas	134
Table 5.6: Results of Vehicle Delay Calculation in Oversaturated Conditions Based on Different Procedures Presented in HCM (2000).....	137
Table 5.7: Mode Choice Attributes and their Coefficients (Parsons et al., 1992; KPMG, 1992; NCHRP 1998).....	141
Table 6.1: Truth Values Table for Fuzzy Logic Membership Values at their Extreme and Truth Values for Boolean Logic	177
Table 6.2: Fuzzy Operators AND, OR and NOT.....	178
Table 6.3: Parameter Values of Membership Functions of the Error, e	189
Table 6.4: Parameter Values of Membership Functions of the Change of Error, Δe	189
Table 6.5: Parameter Values of Membership Functions for the Output Linguistic Value of Green Time Change	191
Table 6.6: Membership Function Values.....	192
Table 6.7: Normalized Parameter Values of Membership Functions for Error, e.....	193
Table 6.8: Various Descriptions of Linguistic Values.....	194
Table 6.9: List of Rules for Two Input and One Output Variable.....	195
Table 6.10: Control Rule Table	195
Table 6.11: Rule Base for the Queue Management Example.....	198
Table 6.12: Values Used in the Centre of Gravity Defuzzification Method	201
Table 6.13: Values Used in the Centre Average Defuzzification Method	204

Table 6.14: Scaled Membership Function Parameters for the Input Variable <i>Error</i> , <i>e</i> ...	207
Table 6.15: Scaled and de-normalized Output Membership Function Centres	211
Table 6.16: Output Membership Functions Scaled to Have Higher Sensitivity at Small Outputs and Smaller Sensitivity at Higher Outputs	213
Table 6.17: Output Membership Functions Scaled to Have Lower Sensitivity at Small Outputs and Higher Sensitivity at Higher Outputs	213
Table 8.1: Queue Control Objectives.....	253
Table 8.2: Tuned Parameters and Control Performance for the “Park View” Type Control	257
Table 8.3: Control Performance for the “On-Off” Type Control	258
Table 8.4: Control Performance for the Model Based Algorithm	262
Table 8.5: Normalized and Scaled Membership Function Parameters for the Error <i>e</i> ...	266
Table 8.6: Normalized and Scaled Membership Function Parameters for the Change of Error Δe	266
Table 8.7: Rule Base for Queue Management with Pre-Signal	267
Table 8.8: Rule-Base Described by Linguistic Numeric Variables.....	267
Table 8.9: Tuned Parameters and Control Performance for the Fuzzy Logic Based Control	273
Table 8.10: Arrival Volume (Demand) Distribution	275
Table 8.11: Travel Times Generated by Baseline Scenario and Alternative Set-Ups	277
Table 8.12: The Impact of Queue Management on Mean Vehicle Travel Time.....	278

Table 8.13: Multi-Modal Transportation System Perspective – Impact on Person Travel Time	279
Table 8.14: The Impact of Queue Management on Mean Vehicle Travel Time	279
Table 8.15: Detection Hardware Requirements.....	282
Table 8.16: Simulation Results: Vehicle Travel Times and its Variability	295
Table 8.17: Comparison of Scenarios – Vehicle Travel Time Perspective	296
Table 8.18: Simulation Results: Person Travel Times and its Variability.....	299
Table 8.19: Comparison of Scenarios: Person Travel Time Perspective.....	299
Table 8.20: The Mean and the 85 th Percentile Travel Time	301
Table 8.21: Comparison of Travel Time Savings Expressed by the Mean and the 85 th Percentile Travel Time.....	302
Table 8.22: Impact of Queue Management on the Relative Advantage of Transit Travel Times.....	306
Table 9.1: Tuned Parameters and Control Performance with $Q_T=15$	321
Table 9.2: Tuned Parameters and Control Performance with $Q_T=30$	322
Table 9.3: Tuned Parameters and Control Performance with $Q_T=40$	323
Table 9.4: V/C Ratios at Major Intersections	332
Table 9.5: Arrival Volume Distribution at the Most Upstream External Link.....	334
Table 9.6: The Impact of Queue Management on Mean Vehicle Delay	342
Table 9.7: The Impact of Queue Management on Total Person Travel Time	342

Table 9.8: Comparison of Travel Times at Specific Sections with no Queue Management and with Queue Management	345
Table 9.9: Congestion Level Scenario Definitions	355
Table 10.1: Sample Records of Detector Occupancies and Corresponding NVDZ.....	368
Table 10.2: NVDZ Estimation Based on Individual Detector Occupancies	371
Table 10.3: Membership Function Parameters	390
Table 10.4: Rule Base Definition and the Corresponding Output Values	391
Table 10.5: Test 1 Scenarios.....	395
Table 10.6: Selected MOP for the Base Case Scenario	396
Table 10.7: The Change of Estimation Error as NVDZ Increases	397
Table 10.8: Test 2 Scenarios.....	397
Table 10.9: MOP for Different Detection Coverage	398
Table 10.10: Test 3 Scenarios.....	399
Table 10.11: MoP for Roadways with Different Number of Lanes	400
Table 10.12: Test 4 Scenarios.....	401
Table 10.13: MoP for Estimates based on Different Number of Records	403
Table 10.14: Test 5 Scenarios.....	403
Table 10.15: MoP for Estimates Based on Detector Locations Relative to the Intersection	405
Table 10.16: Test 6 Scenarios.....	406
Table 10.17: MoP for Different Detection Zone Lengths.....	409

Table 10.18: Test 7a Scenarios	410
Table 10.19: MoP for Different Stage Lengths	412
Table 10.20: Test 7b Scenario, Part 1	413
Table 10.21: MoP for Different Cycle Length, Part 1	414
Table 10.22: Test 7b Scenario, Part 2	414
Table 10.23: MoP for Different Cycle Length, Part 2	415
Table 10.24: Membership Function Parameters for Data Sets 13 (Developed for 50/15 Signal Timing) and Data Sets 30 (Developed for 130/60 Signal Timing)	416
Table 10.25: Output Membership Function Values for Data Sets 13 (Developed for 50/15 Signal Timing) and Data Sets 30 (Developed for 130/60 Signal Timing)	418
Table 10.26: Comparison of Mean Travel Times	423
Table 10.27: Comparison of Total Person Travel Times.....	423
Table 11.1: Existing and Proposed Philosophical Framework	429
Table 13.1: Type A Arrival Distribution	463
Table 13.2: Type B Arrival Distribution.....	465
Table 13.3: Type C Arrival Distribution.....	467
Table 13.4: Input Parameters	469
Table 13.5: Impact of Travel Time Changes on Transit Modal Share	474
Table 13.6: Impact of TCTTr on Transit Modal Share.....	475
Table 13.7: Signal Timing Parameters at the Bottleneck Intersection.....	477
Table 13.8: Eastbound Traffic Volumes used for Control Development Purposes.....	477

Table 13.9: Student's t Distribution Percentile Values Used in Significance Test	478
Table 13.10: "Park View" Algorithm Tuning Parameters and MoP – Set 1	479
Table 13.11: "Park View" Algorithm Tuning Parameters and MoP – Set 2	480
Table 13.12: FL Logic Algorithm Tuning Parameters and MoP – Sample Output 1.....	481
Table 13.13: FL Logic Algorithm Tuning Parameters and MoP – Sample Output 2.....	482
Table 13.14: Travel Times for Individual Simulation Runs for the Base Case Scenario	483
Table 13.15: Travel Times for Individual Simulation Runs for the Queue Management without Signal Control Scenario	483
Table 13.16: Travel Times for Individual Simulation Runs for the "Park View" Control Scenario.....	484
Table 13.17: Travel Times for Individual Simulation Runs for the "On-Off" Control Scenario.....	484
Table 13.18: Travel Times for Individual Simulation Runs for the Model Based Control Scenario.....	485
Table 13.19: Travel Times for Individual Simulation Runs for the Fuzzy Logic Control Scenario.....	485
Table 13.20: Statistical Comparison of the Base Case and QMnoSC Scenarios	485
Table 13.21: Statistical Comparison of the Base Case and "Park Vies" Control Scenarios	486
Table 13.22: Statistical Comparison of the Base Case and "On-Off" Control QMnoSC Scenarios	486

Table 13.23: Statistical Comparison of the Base Case and Model Based Control	
Scenarios	486
Table 13.24: Statistical Comparison of the Base Case and Fuzzy Logic Based Control	
Scenarios	486
Table 13.25: Signal Group Sequencing at Carling and Bronson	487
Table 13.26: Approximate Capacity Calculation.....	487
Table 13.27: Signal Group Sequencing at Carling and Preston.....	488
Table 13.28: Approximate Capacity Calculation.....	488
Table 13.29: Signal Group Sequencing at Carling and Booth.....	489
Table 13.30: Vehicle Travel Perspective: Vehicle Travel Time Difference	490
Table 13.31: Person Travel (Bimodal) Perspective: Person Travel Time Difference	491
Table 13.32: Transit Route Frequencies	492
Table 13.33: Transit Lines and Volumes (Hourly Number of Transit Vehicles During the two –hour Morning Peak Period).....	493
Table 13.34: Volume Balancing Summary for the NB Direction	498
Table 13.35: Volume Balancing Summary for the SB Direction	498
Table 13.36: Signal Group Sequencing at Woodroffe and Baseline	502
Table 13.37: V/C calculation for Balanced Volumes	503
Table 13.38: Signal Timing at Woodroffe and Navaho.....	503
Table 13.39: Signal Timing at Woodroffe and Algonquin	504
Table 13.40: V/C Calculation for Balanced Volumes	505

Table 13.41: Signal Timing at Woodroffe and Meadowlands.....	506
Table 13.42: V/C Calculation for Balanced Volumes	506
Table 13.43: Signal Timing at Woodroffe and Norice	507
Table 13.44: Saturation Flow Rate Measured in the Field	511
Table 13.45: Distribution of Arrival Volumes at Various Locations	512
Table 13.46: Comparison of Mean Travel Times for the Car and Transit Modes – Base Case Scenario.....	513
Table 13.47: Summary of Results.....	515
Table 13.48: Detector Locations for Various Scenarios	516
Table 13.49: Detector Parameters in VISSIM	517
Table 13.50: Signal Timing Scenarios.....	517
Table 13.51: Volume Scenarios.....	518
Table 13.52: Membership Function Parameters	519
Table 13.53: Rule Output Values.....	520
Table 13.54: Comparison of Output Results Based on Vehicle Counts and Fuzzy Logic NVDZ Estimation	521

List of Figures

Figure 1.1: Research Framework.....	13
Figure 2.1: Underutilised Green Time at Intersection A due to the Lack of Downstream Queuing Space	22

Figure 2.2: Underutilized Cross Street Green Time at Intersection A due to Intersection Blockage	22
Figure 2.3: Underutilised Green Time due to Intersection Approach Lane Blockage	23
Figure 2.4: The Environmental Impacts of Queues on Surrounding Land Use.....	24
Figure 2.5: Definition of Signal Groups by NEMA	29
Figure 2.6: Definition of Signal Timing Elements	30
Figure 2.7: Ring-Barrier Stage Definition (NEMA Type Phase Representation)	31
Figure 2.8: The Definition of Signal Timing Intervals, Cycle, and Offset.....	32
Figure 3.1: Two-Intersection System.....	51
Figure 3.2: Simplified Analysis and Signal Timing Procedure by Rouphail et al. (1996)	52
Figure 3.3: Elements of the “Park View” Pre-Signal Set-Up.....	59
Figure 3.4: The Framework of Beauchamp et al. (1997) Fuzzy Logic Controller	68
Figure 3.5: Elements of Ramp Metering.....	75
Figure 3.6: Input Data Collection Locations.....	77
Figure 4.1: The Comparison of the Traditional and the Proposed Planning Philosophy .	88
Figure 4.2: The Components of System Delay.....	92
Figure 4.3: Alternative Priorities at an Intersection Approach.....	96
Figure 5.1: The Relationship between Objectives at Various Levels and MoP in Single-Mode Unsaturated and Bimodal Oversaturated Conditions	102
Figure 5.2: Example 1-Single-Mode Operation at Isolated Intersection	106
Figure 5.3: Sensitivity Analysis.....	110

Figure 5.4: Example 2 – Bimodal Operation at Isolated Intersection.....	111
Figure 5.5: Example 3 – Arterial Operation	113
Figure 5.6: Example 4 – Minimization of Pollution	115
Figure 5.7: Uniform Delay.....	118
Figure 5.8: Graphical Representation of Oversaturated Conditions.....	120
Figure 5.9: Simplified Representation of the Overflow Delay	121
Figure 5.10: The Development of Oversaturated Delay Curve (cyc=60 sec, g=30 sec, s=1800 veh/hr, $t_a=15$ min)	123
Figure 5.11: Description of Arrival Distributions A, B, and C.....	129
Figure 5.12: Arrival Distribution B	131
Figure 5.13: Cumulative Arrivals, Departures, and Delay for Arrival Distribution B ...	132
Figure 5.14: Arrival Distribution C	133
Figure 5.15: Cumulative Arrivals, Departures, and Delay for Arrival Distribution C ...	133
Figure 5.16: Alternative HCM 2000 Approaches for the Analysis of Oversaturated Conditions (Source: TRB, 2000)	135
Figure 5.17: The Urban Transportation Modeling System (UTMS)	139
Figure 5.18: The Impact of Transit Travel Time Change on Transit Modal Share while Everything Else Remains the Same	148
Figure 5.19: Iterative Modal Split Calculating Algorithms	149
Figure 5.20: The General Relationship between TMS and TCTTr	151
Figure 5.21: The Expansion of TCTTr from a Short Section to the OD Pair.....	153

Figure 6.1: Simple Queue Management Set-Up	166
Figure 6.2: Block Diagram of Fuzzy Control Mechanism for Queue Management	167
Figure 6.3: Membership Function Examples	171
Figure 6.4: Trapezoid and Triangular Membership Functions	173
Figure 6.5: Graphical Presentation of the Inference Mechanism when the Output is a Fuzzy Set.....	182
Figure 6.6: Graphical Presentation of the Inference Mechanism when the Output is a Function	184
Figure 6.7: Graphical Presentation of the Inference Mechanism when the Output Function is a Singleton Value	185
Figure 6.8: Membership Functions	190
Figure 6.9: Normalized Membership Functions of Error, e	193
Figure 6.10: Graphical Presentation of the Mamdani Inference Process.....	199
Figure 6.11: Fuzzy Output	200
Figure 6.12: Graphical Presentation of the Takagi-Sugeno Inference Process with Singleton Outputs.....	203
Figure 6.13: Description of Tuning Objectives	206
Figure 6.14: Scaling Input Membership Functions.....	208
Figure 6.15: Scaled Output Membership Function Centres.....	212
Figure 6.16: Control Surface with Output Scaling Factor $sf_R=0.5$	214
Figure 6.17: Control Surface with Output Scaling Factor $sf_R=2$	214

Figure 7.1: Queue Management and Queue Jump at Bottleneck Intersection.....	218
Figure 7.2: Queue Management and Queue Jump Upstream from Bottleneck Intersections	222
Figure 7.3: The “Traditional” Analysis Process for Unsaturated Corridors.....	224
Figure 7.4: Traffic Count Distribution in Corridors with Capacity Limitations.....	228
Figure 7.5: Intersection Operation at Capacity	229
Figure 7.6: Downstream Capacity Limitations- Relationship between the Arrival Distribution, Traffic Counts Distribution, and Capacity	230
Figure 7.7: Upstream Capacity Limitations- Relationship between the Arrival Distribution, Traffic Counts Distribution, and Capacity	231
Figure 7.8: Analysis Process for Oversaturated Corridors	233
Figure 8.1: Traditional Intersection Set-Up without Queue Management	240
Figure 8.2: Simple Conversion of the Queuing Regime from FIFO to LIFO	241
Figure 8.3: Queue Management with Pre-Signal.....	244
Figure 8.4: Geometric Set-Up of the Original “Park-View” Pre-Signal	248
Figure 8.5: The Original Pre-Signal Algorithm by Oakes et al., 1995	250
Figure 8.6: Parameters of the “Park View” Type Queue Management.....	251
Figure 8.7: Generalized and Enhanced Queue Relocation Algorithm Originally Described by Oakes et al. (1995)	252
Figure 8.8: “Park View” Type Queue Control Algorithm Output with Tuned Parameters	256

Figure 8.9: “On-Off” Type Algorithm.....	257
Figure 8.10: “On-Off” Type Queue Control Algorithm Output.....	258
Figure 8.11: Required Data for the Model Based Queue Control	259
Figure 8.12: Queue Management Algorithm for the Model Based Control	261
Figure 8.13: Model Based Queue Control Algorithm Output	262
Figure 8.14: Queue Management Algorithm for the Fuzzy Logic Based Control	264
Figure 8.15: Green Time Adjustment Calculation at the Metering Signal Based on Fuzzy Logic	264
Figure 8.16: The Meaning of Positive and Negative Error and Change of Error	265
Figure 8.17: Fuzzy Logic Simulation Model Framework	268
Figure 8.18: The Role of Tuning in the Queue Management Design Process	270
Figure 8.19: Data Flow during the Tuning Process	271
Figure 8.20: Fuzzy Logic Based Queue Control Algorithm Output.....	272
Figure 8.21: Visual Presentation of Travel Time Comparisons of Alternative Controls	278
Figure 8.22: Multi-Modal Transportation System Perspective – Impact on Person- Travel Time	280
Figure 8.23: The Carling Avenue Corridor.....	283
Figure 8.24: Schematic Lane Arrangement for the “Before” Scenario in the Critical Eastbound Direction.....	285
Figure 8.25: Alternative 1 – Partial Queue Jump Without Pre-Signal	289
Figure 8.26: Partial Bus Queue Jump Lane with Physical Barrier	290

Figure 8.27: Bulb-Out Used to Enforce a Partial Queue Jump Lane.....	291
Figure 8.28: Alternative 2 – Partial Queue Jump With Pre-Signal.....	292
Figure 8.29: Alternative 2 – Pre-Signal Set-Up.....	294
Figure 8.30: Definition of Travel Sections for the Carling Avenue Evaluation.....	295
Figure 8.31: Distribution of Travel Times for the “Before” Scenario.....	297
Figure 8.32: Distribution of Travel Times for Scenario A.....	297
Figure 8.33: Distribution of Travel Times for Scenario B.....	298
Figure 8.34: Distribution of Transit Travel Times – “Before” Scenario.....	300
Figure 8.35: Distribution of Transit Travel Times – Scenario A.....	301
Figure 8.36: Distribution of Transit Travel Times – Scenario B.....	301
Figure 8.37: The Impact if Travel Time Variability (Lateness) on Car and Transit Trips	304
Figure 8.38: Late Car Travel Times are Compensated by Early Car Travel Times.....	304
Figure 8.39: Early Buses Wait at Time Points or at the End of the Route.....	305
Figure 9.1: Elements of Upstream Metering.....	309
Figure 9.2: Physical Set-Up of the Proposed Queue Jump with Queue Management ...	310
Figure 9.3: Queue Management with Queue Length Redistribution.....	313
Figure 9.4: Signal Timing Sequence for Intersection M.....	315
Figure 9.5: Controller Algorithm for the Metering Intersection.....	318
Figure 9.6: New Stage Length Calculations Subroutine.....	319
Figure 9.7: Control Output with $Q_T=15$	321

Figure 9.8: Control Output with $Q_T=30$	323
Figure 9.9: Control Output with $Q_T=45$	324
Figure 9.10: The Woodroffe Avenue Corridor	326
Figure 9.11: Schematic Lane Arrangement in the Critical Northbound Direction.....	327
Figure 9.12: Analysis Process for Oversaturated Corridors (Copy of Figure 7.8)	329
Figure 9.13: Phase Sequencing at Woodroffe Avenue and Navaho Drive.....	330
Figure 9.14: Phase Sequencing at Woodroffe Avenue and Algonquin	330
Figure 9.15: Phase Sequencing at Woodroffe Avenue and Meadowlands Drive.....	330
Figure 9.16: Assumed Distribution of Arrivals	334
Figure 9.17: Geometric Set-up for Queue Management.....	340
Figure 9.18: Definition of Travel Sections	346
Figure 9.19: Comparison of Travel Times at Specific Sections without (“Before”) and with (“After) Queue Management	346
Figure 9.20: Comparison of Car Travel Times During the Simulation Period.....	349
Figure 9.21: Comparison of Transit Travel Times During the Simulation Period	352
Figure 9.22: Transit Travel Time Distribution without Queue Management and the 85th Percentile Travel Time.....	354
Figure 9.23: Transit Travel Time Distribution with Queue Management and the 85th Percentile Travel Time.....	354
Figure 9.24: Impacts of the Level of Congestion on Car and Bus Travel Times without Queue Management	356

Figure 9.25: Impacts of the Level of Congestion on Car and Bus Travel Times with Queue Management	356
Figure 9.26: The Impacts of Queue Management on the Overall Mean Travel Time at Various Levels of Congestion.....	357
Figure 9.27: The Impacts of Queue Management on Total Person Delay at Various Levels of Congestion	358
Figure 9.28: Green Time and NVDZ Variation during the Simulation Period without Queue Management	360
Figure 9.29: Green Time and NVDZ Variation during the Simulation Period with Queue Management.....	360
Figure 10.1: Detection Set-up for NVDZ Estimation.....	367
Figure 10.2: The Relationship between Detector A Occupancy and NVDZ.....	369
Figure 10.3: The Relationship between Detector B Occupancy and NVDZ.....	370
Figure 10.4: The Relationship between Detector C Occupancy and NVDZ Length.....	370
Figure 10.5: The Relationship between all three Detector Occupancies and NVDZ	372
Figure 10.6: Architecture of a Simple Artificial Neuro-Network.....	376
Figure 10.7: The Structure and Functions of an Artificial Neuron/Node	377
Figure 10.8: Sigmoid Function (Logistic Curve).....	378
Figure 10.9: Output Calculation for Node 1 in the Hidden Layer	380
Figure 10.10: Output Calculation for Node 2 in the Hidden Layer	380
Figure 10.11: Output Calculation for Node 1 in the Output Layer.....	381

Figure 10.12: Output Calculation for Node 2 in the Output Layer.....	382
Figure 10.13: The FIS Model Structure Defined in ANFIS	389
Figure 10.14: The Error for the Training and Checking Data	390
Figure 10.15: Steps for Testing Fuzzy Logic NVDZ Estimation	393
Figure 10.16: General Geometric Set up	394
Figure 10.17: Estimated vs. Measured NVDZ for the BCS	396
Figure 10.18: Comparison of the NVDZ Estimates for Different Detection Coverage .	398
Figure 10.19: Comparison of NVDZ Estimates for Operation on Roadways with Different Number of Lanes.....	400
Figure 10.20: Comparison of NVDZ Estimates based on Different Number of Records	402
Figure 10.21: Scenarios for Detector Locations	404
Figure 10.22: Comparison of NVDZ Estimates for Different Data Collection Point Locations.....	405
Figure 10.23: Detection Zone Length Scenarios	407
Figure 10.24: Comparison of Queue Length Estimates for Different Detection Zone Lengths.....	408
Figure 10.25: Comparison of NVDZ Estimates for Different Stage Lengths	411
Figure 10.26: The Application of FIS Developed for Signal Timing 50/15 to Signal Timing 130/60.....	413

Figure 10.27: The Application of FIS Developed for Signal Timing 130/60 to Signal Timing 50/15.....	415
Figure 10.28: FIS Parameter Tuning for Estimating the NVDZ	421
Figure 10.29: Queue Management Controller Algorithm Based on FL NVDZ Estimation	422
Figure 10.30: Comparison of Travel Times for the two NVDZ Calculation Methods...	424
Figure 10.31: Green Time and NVDZ Variation when the NVDZ is Determined by Using Vehicle Counts.....	426
Figure 10.32: Green Time and NVDZ Variation when the NVDZ is Determined by Fuzzy Logic Estimation Based on Occupancy Data.....	426
Figure 13.1: Type A Arrival Distribution: Arrivals and Queue Length	464
Figure 13.2: Type A Arrival Distribution: Cumulative Arrivals and Departures.....	464
Figure 13.3: Type B Arrival Distribution: Arrivals and Queue Length	466
Figure 13.4: Type B Arrival Distribution: Cumulative Arrivals and Departures	466
Figure 13.5: Type C Arrival Distribution: Arrivals and Queue Length	468
Figure 13.6: Type C Arrival Distribution: Cumulative Arrivals and Departures	468
Figure 13.7: “Park View” Algorithm Tuning, Output for Set 1	479
Figure 13.8: “Park View” Algorithm Tuning, Output for Set 2	480
Figure 13.9: Algorithm Tuning, Sample Output for Set 1.....	481
Figure 13.10: FL Algorithm Tuning, Sample Output for Set 2.....	482
Figure 13.11: Timing Sequence at Carling and Bronson.....	487

Figure 13.12: Timing Sequence at Carling and Preston	488
Figure 13.13: Timing Sequence at Carling and Booth	489
Figure 13.14: Transit Frequencies by Transit Lines	494
Figure 13.15: The Volume Balancing Process	497
Figure 13.16: Traffic Counts and Turning Movement Percentages	499
Figure 13.17: Balanced Volumes Through Steps 1 to 8 and Step 9	500
Figure 13.18: Manual Optimization of Signal Splits	501
Figure 13.19: Timing Sequence at Woodroffe and Baseline	502
Figure 13.20: Timing Sequence at Woodroffe and Navaho	503
Figure 13.21: Timing Sequence at Woodroffe and Algonquin.....	504
Figure 13.22: Implemented Timing Sequencing at Woodroffe and Meadowlands.....	505
Figure 13.23: Simulated Timing Sequencing at Woodroffe and Meadowlands.....	505
Figure 13.24: Timing Sequence at Woodroffe and Norice.....	506
Figure 13.25: Impact of driver behaviour on the saturation flow rate	510
Figure 13.26: Distribution of Arrival Volumes at Various Locations	512
Figure 13.27: Detector Set up	517

List of Appendices

Appendix A: Delay Calculation with the Deterministic Queuing Model.....	463
Appendix B: Mean Delay Calculation with the Analytical Formulas	469
Appendix C: Calculation of Transit Modal Share for different Transit Travel Time Changes.....	474

Appendix D: Calculation of Transit Modal Share for different TCTTr	475
Appendix E: Trapezoid Area Calculation.....	476
Appendix F: Development of “Partial” Queue Jump	477
Appendix G: Tests of Hypotheses and Significance.....	478
Appendix H: “Park View” Type Algorithm Development.....	479
Appendix I: Fuzzy Logic Tuning for Pre-Signal Operation Control.....	481
Appendix J: Simulation Results of Control Algorithm for Pre-Signal Operation.....	483
Appendix K: Background Data for the Carling Avenue Example	487
Appendix L: Carling Avenue Evaluation	490
Appendix M: Transit Volumes in the Woodroffe Avenue Corridor.....	492
Appendix N: Volume Balancing.....	495
Appendix O: Optimization of Signal Timing in the Woodroffe Corridor	501
Appendix P: Signal Timing Parameters and the V/C Calculation.....	502
Appendix Q: Saturation Flow Calibration in VISSIM	508
Appendix R: Traffic Count Profiles in the Woodroffe Corridor	512
Appendix S: Simulation Results of the Woodroffe Corridor.....	513
Appendix T: Simulation Parameters for Fuzzy Logic NVDZ Estimation.....	514

Glossary of Terms

actuation - a triggering mechanism activated by the presence of a vehicle or pedestrian

all red interval - the period when signal heads for conflicting movements display red simultaneously (s)

arrival flow rate - the mean number of vehicles arriving (veh/hr, veh/cyc)

bimodal transportation - system consisting of two modes of transportation (e.g. car and transit in this thesis)

bivalent logic - or two valued logic, where the statement or proposition can have either 'true' or 'false' value

bottleneck - see capacity bottleneck

capacity - maximum reasonable number of vehicles or persons that can cross a predetermined point or line during a predefined observation period (veh/hr, veh/cyc, pers/hr, pers/cyc)

capacity bottleneck - elements of the transportation infrastructure with a V/C ratio approaching one, usually applies to intersection approaches but may apply to roadway sections as well

cycle - time during which all signal stages are displayed

check-in detector - vehicle detection sensor at the beginning of the detection zone

check-out detector - vehicle detection sensor at the end of the detection zone

compensating stage - signal stage that gets the green time which has been taken from the metered stage

congestion - the impact of the impedance that vehicles impose on each other in traffic conditions where demand approaches or exceeds capacity

D/C - demand over capacity ratio

degree of saturation – the ratio of volume and capacity (same as the V/C ratio)

detection zone – a section of roadway where the number of vehicles is monitored (counted) and used as an input into the queue management algorithm

detector occupancy - percent of time a detector senses the presence of a vehicle within a specified time period, usually one second

demand – theoretical number of arrivals per unit of time assuming there are no capacity constraints (veh/hr or pers/hr)

density – number of vehicles within a specified length of a lane segment (veh/km)

enforcement by design – implementing design elements that make unintended behaviour obvious and unattractive

fuzzy logic – is a multi valued logic used for reasoning that is approximate rather than precise

headway – time between two vehicles (sec)

heuristic – a method to solve a problem based on educated guesses and intuitive judgement

ideal travel time – travel time from A to B at the speed limit without delay at intersections (sec)

lagging left - left turn signal displayed after the straight through signal in the opposing detection

leading left – left turn signal displayed before the straight through signal in the opposing detection

macroscopic model – traffic simulation model that describes traffic at a high level of aggregation

Measures of Performance (MoP) – quantities used to describe the effectiveness of operation or design (e.g. vehicle delay, V/C, travel time, transit modal share)

mean – measure of central tendency, the arithmetic mean is assumed

metering intersection – an intersection that restricts/meters traffic flow for a specific direction with the objective to control downstream queue lengths

metering stage – signal stage which is modified from cycle to cycle with the objective to control downstream queue lengths

mesoscopic model – traffic simulation model that describe the operation on a more aggregate level than microscopic models but at a less aggregate level than macroscopic models

microscopic model – traffic simulation model that describes the movement of individual vehicles and their interactions

multimodal transportation – transportation system where the characteristics of various modes of travel (e.g. car, bus, cycling, walking) are explicitly recognized

multivalent logic – or multi valued logic, where a statement or proposition can have any value between the true and false values

NEMA – National Electrical Manufacturers Association

NEMA phase – see signal group

NVDZ – the number of vehicles in the detection zone: the vehicle count within a specified length of roadway (detection zone)

offset – the time difference between the start of the green indication at one intersection related to the system time reference point

oversaturated condition – traffic condition in which the arrival flow rate exceeds capacity and the built up queues impact the operation of upstream intersection(s)

person delay – difference between the ideal person travel time and the actual person travel time (sec)

phase – see stage and signal group

primary queue - the queuing vehicles at the particular intersection do not interfere with the operation of the upstream intersection

queue length – the number of vehicles or length of space occupied by vehicles traveling at a very low speed (e.g. $v \leq 5$ km/hr) or stopped

restricted link – same as restricted zone, roadway section between the metering and bottleneck intersections where the number of vehicles is managed

right-turn-on-red – a right turning movement at a signalized intersection that is executed while the red signal is displayed

saturated condition - traffic condition in which the arrival flow rate is close or exceeds

capacity but where the impacts of vehicle queuing are localized to that

intersection (no impact on the operation of upstream intersections)

saturation flow rate – the equivalent hourly rate at which previously queued vehicles

can traverse an intersection approach under prevailing conditions, assuming that

the green signal is available at all times (veh/hr/lane, veh/sec/lane)

secondary queue – queues that are long enough to interfere with the operation of

upstream intersections

signal cycle – one complete sequence of signal indications (stages)

signal group - is the smallest logical unit in signal control, it includes signal heads

always showing the same colors at the same time, one or more signal groups form

a signal 'stage', the group indicates that there could be more displays showing

exactly the same colors, in North America the traditional name for signal group is

NEMA phase

split – the percentage of cycle length allocated to a specific stage in the cycle

stage – represents signal groups that are displayed at the same time, it is also referred to

as 'phase' although this definition is confusing because the 'NEMA phase' has a

different meaning

steady state – a condition of physical system that does not change over time, the system

is in equilibrium

stop bar – white painted line on the roadway surface indicating the position where to stop during the red signal at an intersection approach

transit modal share – the ratio of the number of people travelling by transit and the total number of people travelling by cars and transit

unsaturated – traffic situations where the arrival flow rate is lower than capacity

V/C ratio– see degree of saturation

vehicle delay - difference between the ideal travel time of a vehicle and the actual travel time (sec/veh)

vehicle detector – a sensor usually embedded in the roadway or mounted above the roadway for the purpose of registering the presence of a vehicle

volume – the number of vehicles that cross a specific line per unit of time (veh/hr)

Symbols

C – capacity (veh/hr)

cap – cycle capacity (veh/cycle)

cl – clearance interval (sec)

cyc – cycle length (sec)

e - error

g – green time (sec)

g^{\min} – minimum green time (sec)

g^{\max} – maximum green time (sec)

MoP – Measures of Performance

NVDZ – number of vehicles in detection zone

q - arrival flow rate (veh/sec)

Q_T - target queue length (veh or m)

Q_M – measured queue length (veh or meter)

s – saturation flow rate (veh/hr)

s' – saturation flow rate (veh/sec)

sg – signal group

TMS – transit modal share

TCTTr – transit-car travel time ratio

V - volume (veh/hr)

X - degree of saturation

1. INTRODUCTION

1.1 Overview

Demand for vehicle capacity in most growing urban areas is increasing faster than the society's ability and willingness to supply sufficient roadway space. This trend can be observed not only in older European cities that were originally built for pedestrians and carts with horses but also in North American cities and suburbs that were planned and built from the very beginning almost exclusively for car travel.

The inability to provide sufficient vehicular capacity results in congestion. Although congestion is generally viewed as an undesirable phenomenon, it could also be an indication of success due to the desirability of the roadway infrastructure or the attractiveness of the areas that are served by the roadway. Congestion eventually balances supply and demand in any transportation system. As the level of congestion increases, the attractiveness of the facility or area decreases, thus equilibrium develops. Interest in addressing congestion starts to develop when the society becomes overwhelmingly unsatisfied with the equilibrium. For example, when further growth is desired but prevented by congestion, or when the environmental or social impacts at the equilibrium stage are unacceptable, society demands action. This is the situation which is addressed in this thesis.

Approaches to deal with congestion can focus on the demand and the supply side.

Initiatives on the demand side include efforts from reducing the need for motorised travel

through less car oriented urban design to policy measures aiming to curtail vehicular travel by increasing the cost of travel, by increasing vehicle occupancy through various incentives, redirecting demand to other more effective transportation modes (transit), and other Transportation Demand Management (TDM) measures.

Initiatives on the supply side include building new infrastructure and better utilising existing infrastructure. New infrastructure, in terms of additional roadway and especially transit capacity has an important role. However, transit capacity improvements are generally big scale and high cost projects, thus their implementation is extremely slow.

In built up and growing urban areas, the ability to provide sufficient vehicle capacity along signalised urban arterial is limited. The law of diminishing marginal return in terms of capacity benefits of additional lanes applies to signalised intersections and thus to arterial networks (Mucsi et al., 2002). While the marginal cost of each additional lane increases as the intersection becomes larger, the marginal capacity benefit of each additional lane decreases. As a result, supply cannot match demand, the unavoidable result is congestion. As Goodwin (1997) described it: "...the consequence (congestion) is a matter of arithmetic, not politics...".

Assuming that vehicular congestion cannot be eliminated, and there are strong indications that this assumption is correct, two questions arise:

- What is better, managed or unmanaged congestion?
- What is better, vehicular congestion with or without alternative modes of travel?

The answer to both questions seems to be obvious, in theory. In practice, however, the contrary can be observed. For example, although it is known that the potential capacity increase through signal timing, bottleneck removals on most existing congested urban networks is relatively minor; this reserve capacity is very rarely used for transit operations improvements. Instead, it is regularly used for localized improvements to car travel with limited benefits and with disbenefits to most non-car modes.

The reasons why the mainstream transportation profession has not yet explored the potentials of bimodal congestion management are listed below:

- There is an unrealistic expectation that sufficient vehicular capacity can be built and the denial of physical and financial constraints. As a result, congestion is viewed as a fixable temporary condition which could be resolved from the technical perspective and only requires sufficient financial investment and a political resolve.
- There is a belief that transit modal share is mostly fixed, and expecting a significant increase in transit modal share is unrealistic. This belief is reflected in design and evaluation procedures as well. For example, there is no MoP in the field of Transportation Engineering which would quantify the relative advantage or disadvantage of transit travel compared to car travel. This is despite the fact, that the attractiveness of transit travel (and transit modal share) is significantly impacted by the relative advantage/disadvantage of transit travel times.
- The bimodal optimization of traffic control in oversaturated conditions is more complex than the optimization for single mode operation. Traffic control in single

mode oversaturated conditions is complex enough in itself, adding another layer of complexity by including long-term transit objectives makes it overwhelming.

- There is an expectation that other approaches such as major changes in development patterns, congestion pricing, significant amount of new transit infrastructure will resolve the problem of urban congestion.

The last two reasons why the transportation profession tends to stay away from bimodal transportation management have some merits. However, there are a number of good reasons why exploring the subject is worthwhile:

- Unmanaged congestion decreases the nominal vehicular capacity of the existing transportation infrastructure. On signalised arterials and networks, excess demand produces long and unstable queues which reduce the discharge rate at intersections, block lanes and upstream intersections. By managing congestion, the nominal capacity of the infrastructure can be preserved.
- Congestion impacts not only the users of transportation facilities, but the surrounding environment as well. The transportation network often has capacity bottlenecks which cannot be removed due to various reasons. These reasons often include the undesirable impact on the surrounding land use and space and/or financial constraints. These bottleneck locations are often in areas that have intensive land use. Therefore, from the social and environmental perspective, it makes sense to manage congestion and move congestion to areas where it causes the least damage. Research and implementation initiatives in this area are limited.

- In unmanaged congestion it is impossible to effectively promote higher efficient transportation modes, such as transit. It is known that the relative speed of transit travel (compared to auto travel) is an important factor in determining transit ridership and modal share. In unmanaged congestion, where transit vehicles have the same low and unreliable travel speeds as other vehicles, transit operation is very inefficient, capital and operating costs are high and ridership and revenue are low. Therefore, traffic control and congestion management techniques that provide operational priority to transit vehicles improve the effectiveness and efficiency of the transportation system.
- Although changes to past and existing car oriented urban planning paradigms have been initiated and the introduction of road user charges seems to be almost inevitable in the longer-term, the positive impacts generated by these future measures could be many years ahead. On the other hand, bimodal transportation management techniques could be implemented quickly.

1.2 Problem Statement

Transportation management and control techniques for managing congestion and for supporting higher efficiency transportation modes such as transit in a mixed flow environment are needed in most urban areas. However, the prevailing traditional transportation planning/management philosophy, traffic control techniques, and evaluation framework do not support the bimodal congestion management concept.

Most existing traffic control techniques are based on mathematical traffic delay models that were developed primarily for uncongested conditions and they do not make distinction between lower and higher efficiency modes. The primary optimization objective of these models is to minimize vehicle delay. However, these models do not differentiate between low or high capacity vehicles and cannot predict the impacts of queue blockages, the main source of delay in congested conditions. Thus, existing vehicle delay based models are ineffective for optimizing bimodal operation in congested conditions.

Urban transportation management has both short-term operational and long-term planning objectives. Short-term operational objectives usually include the maximization of infrastructure capacity and the minimization of delay. On the other hand, long-term planning objectives (in most Canadian cities) normally include significant increase of transit modal share. Traditional traffic control practices focus only on a single mode and only on short-term benefits and they do not support long-term planning objectives. This contradiction however remains hidden when the existing evaluation/analysis framework is utilized.

1.3 Research Objective and Scope

The main objective of this research is to develop a new bimodal traffic control and queue management approach that will allow better utilisation of the existing infrastructure.

More specifically, a traffic management strategy and on-line traffic control algorithms

will be developed that preserve the nominal vehicular capacity of a corridor, decrease total person travel time, and improve the relative travel time of transit vehicles compared to cars.

The second objective is to develop a performance evaluation framework and performance measures that are relevant to the evaluation of bimodal traffic control strategies and measures, and which can facilitate the connection between short-term operational objectives and long-term planning objectives.

The third objective is to identify the required philosophical framework that is needed for the development of the bimodal concept.

In terms of scope, the proposed control techniques focus on the operation of capacity bottlenecks. However, the proposed philosophical approach and evaluation framework applies to any corridor or network.

The proposed traffic control and queue management approach is not a stand alone method. It is built on the elements of traditional signal control by adding new queue control and management functionality. The new functions apply to saturated and oversaturated conditions without any impacts on the operation in unsaturated conditions.

1.3.1 Why is this Research Different from Others? What is the Contribution to Knowledge?

Quite a few researchers have already investigated the topic of traffic management and control in congested conditions. However, the approach suggested in this thesis is different from any other published research:

- It looks at congestion from the holistic perspective
- It has a bimodal focus
- The focus is on signalised arterials
- It combines heuristics with existing control models
- The focus is on the practicality of the method
- **Holistic perspective to congestion**

Transportation research in the area of traffic congestion/control often focuses on relatively narrowly defined problems. This allows researchers to concentrate on very specific issues with intensive focus. However, the problem of traffic/transit operation/control in congested conditions cannot be effectively resolved if there is no supporting philosophical environment and effective performance measuring framework in place. This thesis deals with both of these supporting issues.

Bimodal focus

Transportation research in the area of traffic congestion/control mostly focuses on single mode operation only. The challenges and potential benefits of including transit vehicles in the research are most of the time ignored.

Focus on signalised arterials

The idea of preserving the nominal vehicular capacity through volume metering is not new. The concept is sometimes called queue or volume management. In most cases, volume metering is used on ramps approaching freeways, hence the name 'freeway ramp metering'. Freeway ramp metering has been implemented since the late 1950's and by 2001 ramp metering systems existed in over twenty metropolitan areas within the United States, along with numerous cities around the world (Pearson, 2003). However, there is much less research and very few documented examples of volume or queue management techniques along signalised arterials or networks.

The term 'metering' is normally used when referring to volume management on freeway ramps. However, in case of signalised arterials or networks, the metering techniques usually get a different name such as *gating*, *queue management*, *bus sluicing* or a name without any indication to the metering process such as *traffic control with pre-signals*. More recently metering techniques such as "gating" in conjunction with bus queue jumps have been applied at a number of locations in London, England (Hounsell, 2005). However, published research on queue management strategies and signal control techniques along signalised arterials are limited. The focus of this research is on volume management techniques for signalised arterials.

Techniques based on heuristics and not on mathematical optimization

Most existing traffic control algorithms are based on mathematical models which can provide a relatively effective control for unsaturated conditions. However, in an

oversaturated environment, the same algorithms are ineffective. Optimization algorithms for unsaturated conditions focus on the minimization of vehicle delay, but the formulation of the optimization problem rarely includes queue blockages which have substantial impact on delay in oversaturated conditions. Significant efforts have been made at the research level to produce control procedures for congested conditions, but the problem is far from being resolved.

In this thesis, a different approach to traffic control is suggested for congested conditions. The motivation for the different approach comes from the traffic officer standing in the middle of a busy congested intersection. He looks at all approaches, compares the queue lengths, estimates the negative impacts of queues, looks at the position of the bus, checks pedestrian crowding at the intersection, and makes a decision on who goes first, second and so on. At the same time, he makes sure that someone is always moving and no capacity is lost. If a traffic officer can provide a relatively effective control technique for oversaturated conditions, is it possible to capture some of the human logic and expand the functionality of existing control algorithms? The proposed approach to traffic management will combine the logic of a knowledgeable human traffic officer (heuristics) with existing control algorithms.

The practicality of the proposed approach

It is expected that the results of the proposed transportation management techniques and evaluation procedures will be directly applicable mostly with relatively minor changes to existing signal controller equipment and within a short time frame.

1.3.2 Limitations of the Proposed Approach

There is no claim made that the proposed traffic control approach will allow congestion free travel for all roadway users. The assumption is that congestion free travel in developed and growing urban areas is often practically impossible. However, there is a claim that the suggested methodology will reduce total person travel time, it will increase transit's modal share and thus it will reduce the number of cars and car passengers impacted by congestion.

The proposed control techniques focus on capacity bottlenecks, but it is expected that the general concept of metering traffic and the evaluation framework can be applied to any transportation system. For example, in congested downtown areas with lots of access points, short block lengths, intensive parking activities, the applicability of the proposed control techniques is likely limited. In these situations external metering (Section 4.3) could be applied. This thesis does not deal with external metering techniques, however, the proposed philosophical approach to congestion control and the evaluation framework would apply.

1.4 Study Methodology

In terms of investigating the philosophical environment that is required for the development of the bimodal transportation concept, a literature review, analysis and discussions is carried out. Furthermore, some of the fundamental technical reasons why queue management is needed and why it is effective are explored on elementary

examples. This philosophical framework then provides the necessary background environment for developing both control algorithms and performance evaluation measures.

The main focus of this thesis is on the development of bimodal queue management/control algorithms (Figure 1.1). The development process starts with the identification of control objectives. These objectives may include the elimination of queue overflows, lane and intersection blockages. At the same time, appropriate actions based on human logic are developed.

The human logic is described by a set of strategic rules that apply to various traffic situations. For example, removing an undesirable queue at a bottleneck intersection can be done by holding traffic back by reducing green time at the upstream intersection(s), by increasing the green time at the downstream bottleneck intersection, or by the combination of the previous two methods. Rule development requires the identification of variables that are used as input and output values in the traffic control/management model.

The next step is the traffic control/management model development. The control/management models are developed by using VISSIM's VisVap signal control algorithm development tool. The control/management model uses elements of traditional signal control, special algorithms, and fuzzy logic control.

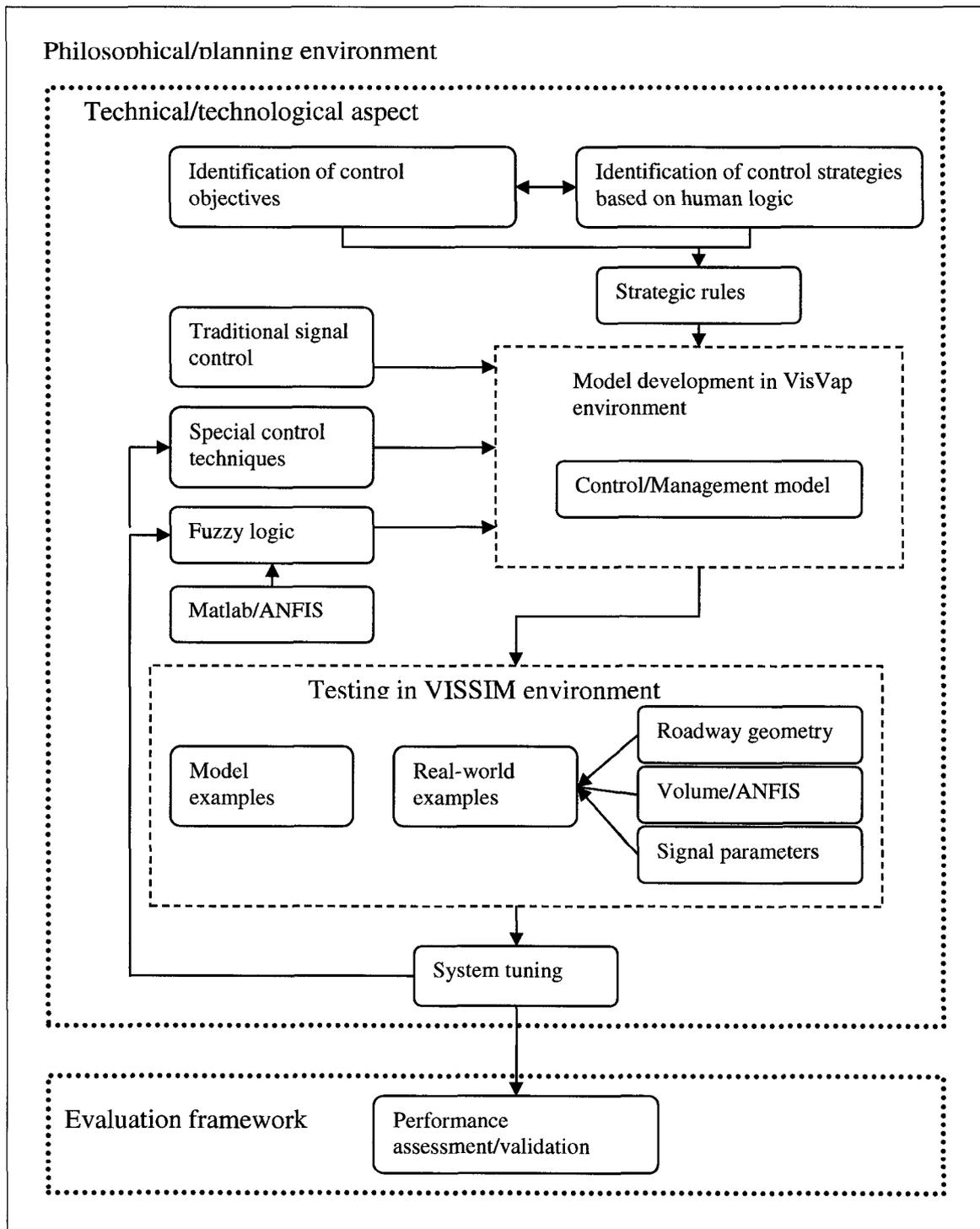


Figure 1.1: Research Framework

The proposed control algorithms can be used in both simulation and real-world environments. In a simulation environment, controller operation can be based on data collected through the microscopic traffic simulation tool. However, in real-world applications, reliable data collection is more difficult. A data collection method utilising an Adaptive Neuro Fuzzy Inference System (ANFIS) is proposed.

The control/management models are tested in a VISSIM environment. VISSIM is a microscopic traffic simulation software which supports externally developed signal control programs (e.g. VisVap programs) and facilitates comprehensive operational performance analysis.

Testing is carried out on model intersections and real-world examples. The model examples are designed to demonstrate the required functionality of the developed control/management techniques. They are also used for system tuning when required and for sensitivity analysis. Real-world examples are taken from Ottawa. Roadway geometry information, required signal timing parameters, and traffic/transit volume and operating characteristics are collected from the City of Ottawa and used in the model. The objective of model testing on real-world examples is to demonstrate the benefits of the bimodal control concept.

The final step is the model performance assessment. Performance data is collected from the VISSIM model. The performance assessment has two objectives: The first objective is to confirm that the control/management model follows the human logic developed at

the beginning of the process. The second objective is to validate the approach by comparing the proposed control/management approach with traditional traffic control practices.

The second objective requires the development of Measures of Performance (MoP) that are appropriate for bimodal operation in congested conditions. To be able to sideline the use of traditional MoP their limitations are analysed and discussed.

1.5 Organization of the Thesis

The thesis consists of eleven chapters. The eleven chapters are grouped into six somewhat distinct parts.

Part A: Introduction and Background Information

Chapter 1 starts with an overview of approaches to tackle urban traffic congestion. This is followed by the problem statement, research objectives, scope, and the limitations of the proposed methodology. The section is completed by describing the research methodology.

Chapter 2 starts with the definition of the most important terms. Then, it provides a brief background on signalised intersection control including the history of signal control, signal control fundamentals, and the most popular signal timing/optimization models and traffic control systems.

Chapter 3 summarizes the literature search. It describes various approaches to traffic management in oversaturated conditions. There is a discussion on the experience of

special queue management techniques including the gating and metering concepts. Since fuzzy logic will be applied in later chapters, there is a literature review on the history of fuzzy logic and its application to traffic control problems.

Part B: Philosophical Background for Bimodal Transportation Management

The implementation of multi-modal transportation management requires a new paradigm in transportation management and control. The objective of **Chapter 4** is to discuss the philosophical background that is required for the paradigm shift. It discusses the failure of the ‘predict and provide’ approach in transportation planning, conceptually describes the consequences of congestion and explores the underlying reasons why the bimodal queue management concept is beneficial.

Part C: Performance Evaluation

The implementation of bimodal transportation management cannot be justified by using the existing traditional performance evaluation framework. **Chapter 5** explores the connection between high level transportation policy/planning objectives and the MoP at the implementation, evaluation, and control levels. It continues with the description of the “gaining efficiency by abusing efficiency paradox” and the resulting “propagation of inefficiencies trap”. Finally, the characteristics of some of the most important traditional MoP are explored and bimodal MoP and evaluation framework are proposed.

Part D: Fuzzy Logic and Queue Management Applications

Chapter 6 explains the concept of Fuzzy Logic (FL) and its implementation to a generic queue management problem. The chapter also describes the proposed tuning method.

Part E: Development and Application of Bimodal Queue Management

Chapter 7 proposes various methods to control the formation of queues, explores the challenges pertaining to the identification of capacity bottlenecks, and proposes a method for the analysis of oversaturated arterials.

Chapter 8 presents the development and evaluation of five different queue management techniques. One of the queue management algorithms is applied to a real-world problem in a simulation environment.

Chapter 9 presents the development of a somewhat different queue management scheme which is applied upstream from the bottleneck intersection. Beside a model example, the chapter includes a real-world example as well. The performance of the proposed control is compared to the performance of the traditional (existing) control.

Part F: Development of Data Collection Technique

Chapter 10 identifies some of the challenges of data collection in real-world applications of the proposed queue management techniques and proposes a new technique, based on fuzzy logic tuned by ANFIS, to collect and process data for the application of queue management.

Part G: Closure

Chapter 11 provides final discussions, conclusions, and proposes future steps for research and implementation.

2. BACKGROUND

2.1 Introduction

The purpose of this chapter is to define the fundamental terms used in the thesis, to demonstrate some of the operational consequences of congestion, and to provide a general background in signal control.

2.2 Definitions

The term ‘congestion’ is used to describe conditions where the interactions between users of a facility reduce their travel speed. Normally this happens when demand for roadway space exceeds the capacity in that particular condition. In many cases, the general description of congestion is too vague and more precise definitions are needed to describe the various levels of congestion.

Longley (1968) differentiates between primary and secondary congestion. *Primary congestion* is caused by the formation of primary queues at capacity bottlenecks such as signal controlled intersections. This condition is often referred to as *saturated condition*.

Pignataro et al. (1978) define saturated conditions as congested conditions where the impact of queues remains localized to the immediate upstream area of the capacity bottleneck. *Secondary congestion* forms at an upstream intersection (secondary queues) due to backups from the primary congestion (Longley, 1968). Oversaturation is another term to describe these conditions (Pignataro et al., 1978). In oversaturated conditions, the formation of queues negatively affects the operation of upstream intersections.

Urbanik (2007) defines oversaturation as a system issue where significant cycle failures occur and traffic signal timing adjustment at the bottleneck intersection cannot improve performance. A cycle failure means that there are unserved vehicles left after the green signal has terminated.

The Highway Capacity Manual (HCM) 2000 defines *unsaturated conditions* when the arrival flow rate is lower than the capacity of the facility (TRB, 2000). A saturated condition applies to conditions when the arrival flow rate is close to capacity, and oversaturation applies to conditions when the arrival flow rate exceeds capacity. The HCM 2000 uses the volume over capacity ratio (V/C) to describe the level of congestion at intersections. However, there are no clear cut-off points between unsaturated, saturated, and oversaturated V/C values.

The HCM 2000 definition of congestion is more precise, since it provides a measure that can be calculated. Longley's definitions of primary and secondary queues, and Pignataro's definition of undersaturation and oversaturation are less precise but they are "richer", since they describes the consequences of oversaturation.

Fortunately there is no contradiction in these definitions of congestion, and all of them are applicable to this thesis. For the purpose of this thesis, the term oversaturation is used to describe conditions where queues impact the operation of upstream intersections. The term saturation is used to describe operation close to capacity. This condition would

normally result in some cycle failures. The term congestion is used when the differentiation between the two levels is not relevant.

2.2.1 Operational Consequences of Congestion

The relationship between travel speed and the number of vehicles on a facility (density expressed as vehicles per units of distance) was one of the first phenomena studied by the traffic engineering profession. As the number of vehicles on a facility increases, due to interactions between the individual vehicles, their travel speed decreases. With only a small number of vehicles on the facility, there is no interaction between vehicles and they travel at a free flow speed. At the other extreme, with a large number of vehicles on the facility, the travel speed is approaching zero.

A number of researchers tried to describe the relationship between speed and density. Roess et al. (1998) provides a good summary of the proposed relationships. Grienshields (1934) proposed a linear relationship, Ellis(1964) developed a piecewise linear model, Greenberg (Roess et al., 1998) introduced a logarithmic relationship, Underwood (1961) suggested an exponential model, Edie (1961) combined Greenberg's equation for high densities and Underwoods equation for low densities, and Duke et al., (1967) experimented with the bell curve.

The formulation of the density – speed relationships has developed over the years, and there is a basic recognition of the fact that when roadways with uninterrupted flow

(without signals or intersections) become overloaded their vehicle carrying capacity drops.

The exploration of the same phenomenon for signalised arterials is much more complex. There is no simple, easily understandable relationship between speed and density for interrupted flow. The analytical description of the impacts of interactions between vehicles on signalized arterials would require many variables describing not only the physical characteristics of traffic flow, roadway geometry, and signal operation but also the variability of human behaviour. A more feasible approach to explore the relationship between speed and the number of vehicles on signalized roadways is through microscopic simulation.

The negative impacts of overloading signalised arterials go beyond the reduction of travel speeds. The impacts can be grouped into four categories:

- Reduced vehicular capacity of the facility
- Environmental impacts due to queue formations at wrong places
- Negative impact on transit
- Negative impacts on pedestrian traffic

2.2.1.1 Reduced vehicular capacity of the facility

The main reason for capacity loss of overloaded signalised arterials and networks is due to lane blockages. Lane blockages prevent the optimum discharge at intersection approaches with the green signal. Figure 2.1 depicts a situation where due to long queues

developed at a bottleneck downstream intersection or slowly moving queues prevent the discharge at upstream intersection 'A' at the saturation flow rate.

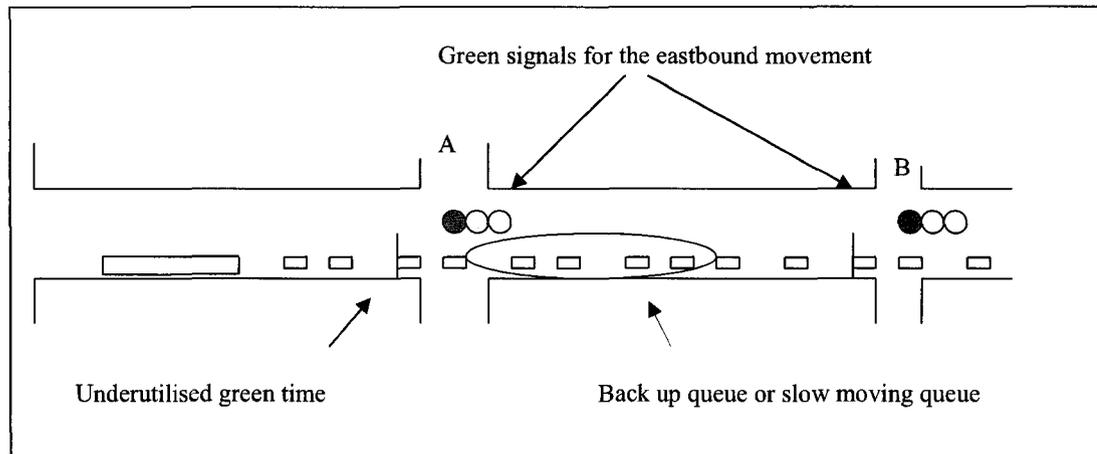


Figure 2.1: Underutilised Green Time at Intersection A due to the Lack of Downstream Queuing Space

Figure 2.2 illustrates an example where left over queues in the intersection prevent the maximum utilisation of cross street green times.

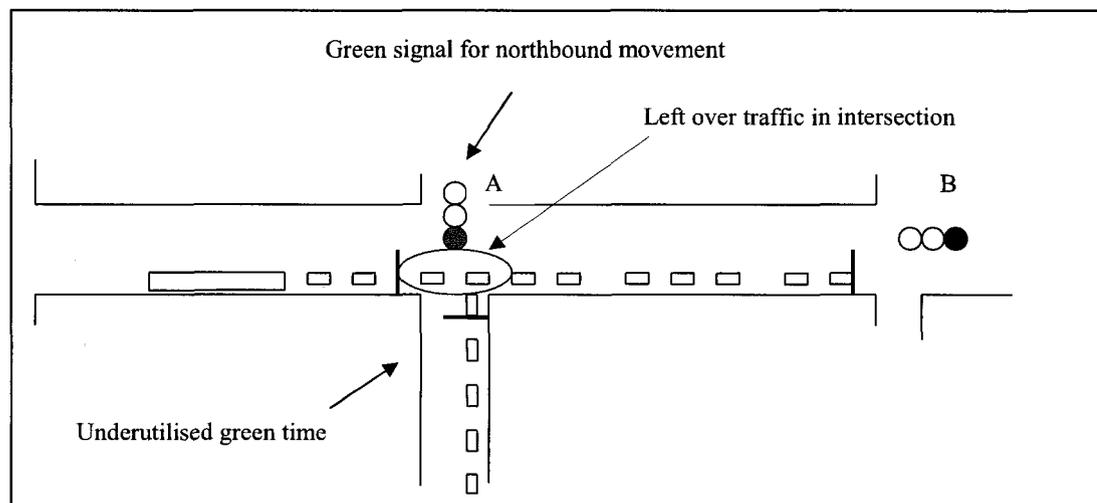


Figure 2.2: Underutilized Cross Street Green Time at Intersection A due to Intersection Blockage

Figure 2.3 shows two examples where vehicles waiting for their green phase prevent other vehicles to utilize a different green phase.

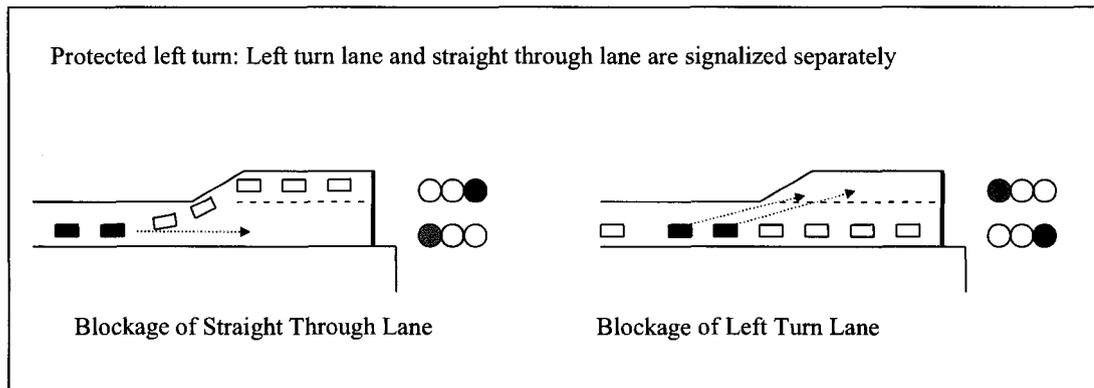


Figure 2.3: Underutilised Green Time due to Intersection Approach Lane Blockage

The Traffic Engineering community has developed approaches to deal with some of the above operational problems. These approaches include driver education and legislative measures that attempt to discourage vehicles from entering the intersection unless they can clear it during the same green phase. This approach, if enforced, is effective in preventing blockages of cross streets as described in Figure 2.2 but it will not prevent the underutilisation of the green phases described in Figure 2.1. The blockage of straight through or left turn lanes as described in Figure 2.3 can often be mitigated by signal phasing changes and geometrical modifications such as left turn lane extensions.

Obviously, the left turning lane extension has its limits, and the signal phasing changes will lose their effectiveness as volume increase beyond the capacity of the intersection.

The consequences of the above lane blockages are severe in terms of capacity. When they occur, the capacity of the arterial and network decrease significantly.

2.2.1.2 Environmental Impacts Due to Queue Formations at Wrong Places

Idling vehicles in long queues have a negative impact on the surrounding environment. Bottlenecks often develop in parts of urban areas that have lots of roadside activity. It is preferable and sometimes possible to move vehicular queues to areas with less roadside activity and still utilise the maximum capacity at the bottleneck intersection. Figure 2.4 describes such a scenario.

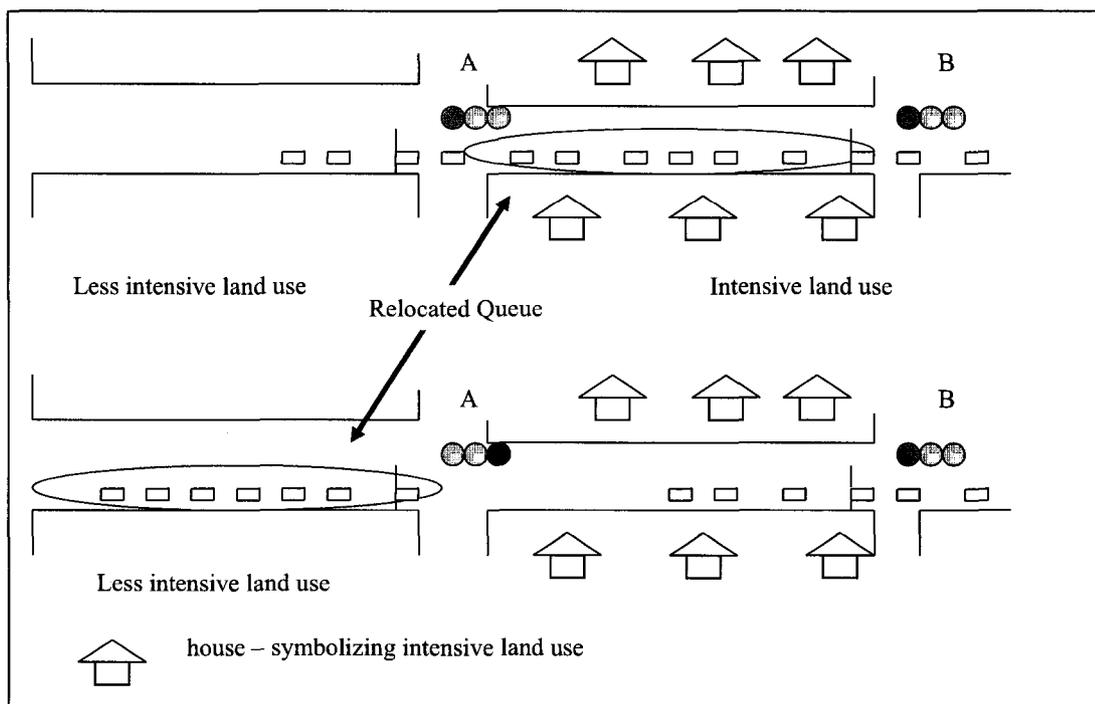


Figure 2.4: The Environmental Impacts of Queues on Surrounding Land Use

2.2.1.3 Negative Impacts on Transit Service

Without special provisions for transit vehicles, the impact of congestion on transit operation and transit ridership is severe. The inability of transit vehicles to clear

intersections during the first green phase creates significant schedule adherence problems. For example, at an intersection that operates at a 120 second cycle length, the inability of the transit vehicle to clear the intersection during the first green phase generates at least one minute of delay. This delay is then propagated through the rest of the route. The consequences of poor schedule adherence include high operating and capital costs of operating the transit service and revenue losses due to the unattractive service.

2.2.1.4 Negative Impacts on Pedestrian Traffic

Intersection blockage described in Figure 2.1 encourages unsafe pedestrian behaviour. Pedestrians waiting to cross in the North-South direction often will start crossing against the red light since the conflicting vehicular queue is not moving. It can be argued that this type of unsafe pedestrian behaviour should be dealt with by enforcement. However, a different approach can be taken by asking the following question: Why is the green phase provided for vehicular movement that cannot proceed? Why isn't the green phase provided to users that could fully utilize the available green time?

Figure 2.2 shows a slightly different scenario. North-South pedestrians have a green light but the pedestrian crosswalk is blocked by vehicles forcing pedestrians to make inconvenient and potentially unsafe movements.

2.3 Traffic Signal Control: A Brief History

The first known installation of a signal device for regulating vehicles, horse riders and pedestrians was invented by JP Knight, a railroad engineer, and installed in 1868 in London, England (Mueller, 1970). The signal was operated by a police officer and had an arm which, when extended horizontally, meant “stop” and when lowered to a 45 degree angle, meant “caution”. The signal was installed near the House of Parliament and its purpose was to prevent collisions between pedestrians, mostly Members of the Parliament and other users (Mueller, 1970).

In 1912, a short lived attempted was made in Paris to control traffic with signals. A kiosk was erected in the middle of an intersection and a four-sided box on the top painted red and white for stop and go was rotated manually. Apparently taxi drivers entertained onlookers by ignoring the signals, and the scheme was discontinued after 22 days (Mueller, 1970).

Probably the first electric traffic signal installation in North America was in Salt Lake City, Utah in 1912, with little publicity. The manually operated signal with red and green displays was designed by Lester Wire, a police officer and was manually operated (Session, 1971).

The installation in Cleveland, Ohio in 1914 was much better publicized (Motorist, 1914) and it is often regarded as the first electric traffic signal installation in North America (FHA, 1995). The American Traffic Signal Company installed red and green lights at

each corner of an intersection. Similar to previous installations, the signal was operated by a police officer but it also had an electric timer (Mueller, 1970).

The three color traffic signal with red, yellow, and green displays, very similar to the operation today, was implemented by William Potts, a police superintendent, in Detroit in 1920. This installation was followed very shortly by installations at four adjacent intersections. The operation of signals at all five intersections was controlled manually by a policeman from a single tower. Controlling five locations from a centrally located tower, however, proved to be problematic, and in 1921 Potts installed an automatic signal timing system at 15 intersections (Motor News, 1947).

Approximately at the same time, John Harriss, Special Deputy Police Commissioner in charge of traffic control, implemented a somewhat different three colour traffic signal along 5th Avenue in Manhattan. Although red, yellow, and green signals were used to control traffic, the display of signals differed from Detroit's. For 5th Avenue, only a red and yellow signal was shown, and for side streets only green and yellow signals were displayed (New York Times, February 6, 1920; New York Times March 15, 1920). After experimentations, a coordinated system was developed by switching all signals manually from one central location, the Time Square. For 30 to 45 sec traffic was allowed to travel along 5th Avenue at all signals and after that cross street traffic proceeded for 30 sec (New York Times, January 5, 1922).

By the 1950s, electromechanical pre-timed devices became the dominant controller hardware in most North American cities. During the 1950's and 1960's many cities started to develop coordinated systems (Urban Traffic Control Systems -UTCS) by interconnecting electromechanical controllers. Coordination among intersections was achieved by operating signals at a common cycle length and by staggering the start of green phases to allow vehicle progression (Black, 1998).

Over the next decades, the development of electronic technology, digital computers, microprocessors, integrated circuits, and communication technologies allowed the replacement of the old electromechanical controllers with relatively inexpensive, reliable solid state controllers, and centrally operated and/or monitored control systems.

2.4 Traffic Signal Operation

The original objective of signal control was to eliminate or minimize the simultaneous movement of conflicting flows through an intersection. As adjacent intersections became signalized, traffic signal control got a second objective in the form of the coordination of movements between adjacent intersections.

To automate this control process, a number of parameters have to be defined. The main timing parameters describing traffic signal operation at an isolated intersection are the stage length and the cycle length. If it is desired to coordinate signals at two or more intersections, a third parameter, the offset has to be defined as well.

The signal indications that allow individual movements (e.g. northbound left turn, or northbound straight through movement) are defined as ‘signal groups’ Figure 2.5 shows signal groups 2, 4, 6, and 8 for the northbound (NB), eastbound (EB), southbound (SB), and westbound (WB) movements. Signal groups 1, 3, 5, and 7 represent the southbound left turn (SB LT), westbound left turn (WB LT), northbound left turn (NB LT) and the eastbound left turn (EB LT). The numbering scheme follows the NEMA (National Electrical Manufacturers Association) recommendation but it can be rotated if required.

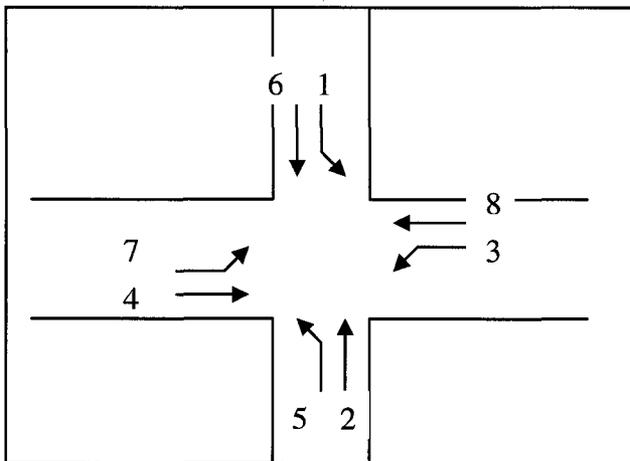


Figure 2.5: Definition of Signal Groups by NEMA

A ‘signal stage’ consists of ‘signal groups’ that allow vehicles to enter the intersection at the same time. Figure 2.6 and Figure 2.7 represent possible staging diagram. The stage is also sometimes referred to as the ‘split’.

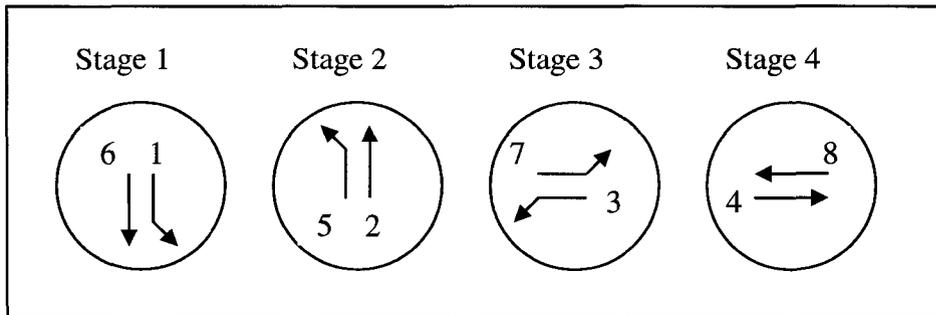


Figure 2.6: Definition of Signal Timing Elements

The terms ‘signal group’, and ‘signal stage’ are uncommon in North America.

In the North American traffic engineering jargon, the term ‘signal group’ is called a ‘signal phase’ or ‘NEMA phase’. However, the term ‘signal stage’ is also referred to as a ‘signal phase’. Differentiating between these two similar but fundamentally different terms is essential in this thesis. To minimize confusion the terms ‘signal group’ and ‘signal stage’ are used exclusively in the rest of the thesis.

The ‘cycle’ length represents the total time required for a complete sequence of signal stages. If a stage is displayed in every cycle regardless of demand, it is called non-actuated stage otherwise it is called actuated stage. For a stage to be actuated, vehicle sensors (or detectors) must be installed at the intersection approach. Actuated stages are not only displayed when there are vehicles on the detectors, but they are also terminated when there no vehicles present, or they are extended when there is a stream of vehicles entering the intersection during the green interval.

Figure 2.7 depicts a more structured way of describing signal groups and stages. This is often called the “NEMA phasing diagram”. The diagram depicts all 8 stages grouped into

two rings. Ring A contains signal groups 1, 2, 3, and 4 and Ring B contains signal groups 5, 6, 7, and 8. Signal groups in the two rings are timed down to a certain degree independently. For example signal group 1 can switch to signal group 2 independently from signal group 5 switching to signal group 6. However, signal groups in the two rings must cross the barriers at the same time, i.e. signal group 2 and 6 must terminate at the same time to display signal groups 3 and 8. This practically means that signal group 1 can be displayed at the same time with signal groups 5 or 6, but it cannot be displayed with signal group 2 at the same time. The width of the rectangular represents the duration of the green plus clearance interval of the signal group.

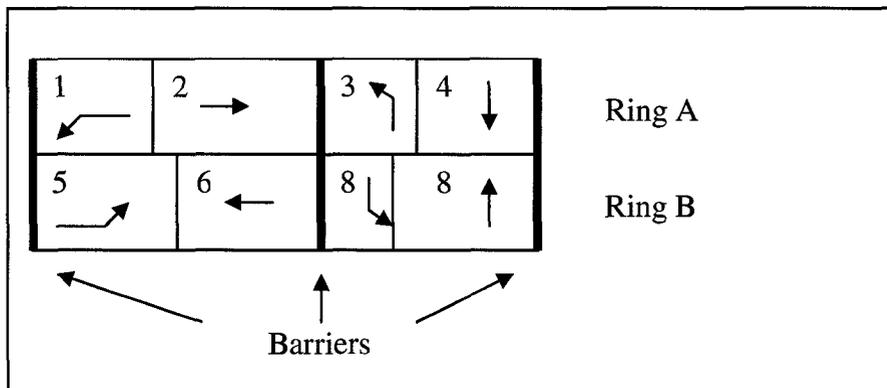


Figure 2.7: Ring-Barrier Stage Definition (NEMA Type Phase Representation)

Signal stages consist of intervals (Figure 2.8). There are four vehicle intervals, the green, amber, all red, and the red intervals. The amber and all-red intervals represent the clearance interval. The purpose of the clearance interval is to prevent collisions of conflicting movements when signal indications change. Pedestrian signal groups comprise the walk, flashing don't walk (FDW), and don't walk intervals.

The cycle length can be defined in various ways, but the differences are not pertinent to this thesis. Figure 2.8 shows one definition of the cycle length, the time duration between the beginnings of the pedestrian Flashing Don't Walk intervals of the coordinated phase.

Signal operation at adjacent intersections are often coordinated by staggering the starts of the green signals with the objective to generate the least delay or/and number of stops.

The start time of the green interval (or sometimes other intervals) at each intersection along the coordinated direction is referenced to a common clock time and it is defined by the 'offset' (Figure 2.8).

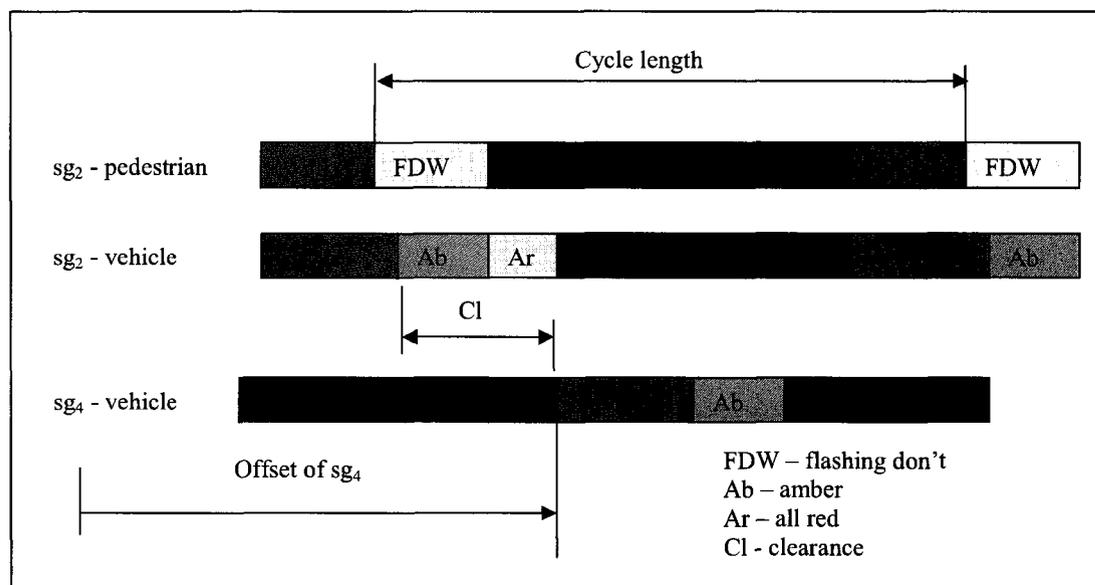


Figure 2.8: The Definition of Signal Timing Intervals, Cycle, and Offset

2.4.1 Signal Timing/Optimization Models

Signal operation is normally designed to minimize average vehicle delay, stops, fuel consumption, or a combination of these factors. The calculation of delay at an

intersection approach is presented in Section 5.3.1. Here, a general description of the signal design procedure and the most common tools are provided.

Signal timing plans are based on historical or projected traffic volumes. Although there is no one unique signal timing design method, most of the following steps are incorporated in every signal timing design method:

- Selection of stages and stage sequencing
- Timing of “yellow” and “all red” intervals for each signal group
- Determination of cycle length
- Allocation of effective green time to the various signal groups/stages
- Checking/calculating pedestrian crossing requirements
- Offset calculation (for arterial and network signal timing)

Normally, a series of timing plans are prepared for different recurring traffic conditions or for different times of day. These plans require data collection and analysis for each situation and/or time of day for which the plan is required. As traffic conditions change, plans need to be updated so that the model could generate an optimum timing plan.

There are two general control approaches, pre-time control and adaptive control. In pre-time control timing plans are changed at specific time determined by historical demand changes. Adaptive control adjusts the timing plans as traffic volumes change. If timing is adapted for each cycle, the approach is called real-time control.

There are a number of traffic/signal operation models/design tools. The most popular in alphabetical order include the Highway Capacity Manual (HCM), Passer, SIDRA, Synchro, and TRANSYT.

HCM

The current version of the Highway Capacity Manual (HCM 2000) is the result of developments in the traffic engineering knowledge for over half a century. It is the most widely used traffic engineering manual in North America. The Highway Capacity Software (HCS) is a tool developed by a third party that uses HCM models in the calculations. The HCM and the HCS have separate sections that deal with signalised intersection control.

The HCS contains an analytical mesoscopic model for signal operation analysis that includes deterministic and stochastic elements. The model focuses on mean vehicle delay and the number of stops. The delay model was developed primarily for isolated unsaturated intersections but by using a progression factor, the model is capable of providing some estimates of coordinated operation as well. However, the accuracy of modeling coordinated operation is an issue subject to further research and development (Prassas, 2002). The HCM methodology has some capability to analyse oversaturated conditions but it was not designed to be used for the calculation of timing plans for those conditions. The model has a vehicle focus, and was not designed to be used for bimodal analysis.

SIDRA (Signalized & unsignalized Intersection Design & Research Aid) is a microanalytical traffic evaluation tool that employs lane-by-lane and vehicle drive cycle models (Akcelik, 2007). Delay and the number of stops are the two basic MoP but other secondary measures, such as fuel consumption, emissions, operating and travel costs can be also included. The lane-by-lane analysis allows the analysis of queue blockage due to the formation of long queues, but the model does not deal directly with oversaturated conditions. The tool does not facilitate separate signal timing design or operation analysis for transit vehicles.

Synchro is an analytical mesoscopic traffic and signal operation and signal timing design tool. It can optimize signal timing based on the linear weighted combination of percentile delay, stops, and queues. In terms of oversaturated conditions, since version 6 Synchro provides some capabilities for the analysis of oversaturated conditions by including queue interaction delay. The queue interaction analysis determines additional delays caused by starvation, spillback, and storage length blocking. Queue delay is an integral part of the optimization objective.

PASSER (Passer II-80, Passer III) is a macroscopic, deterministic, bandwidth optimization tool for the analysis of traffic operation and for the calculation of signal timing plans for arterials. It calculates cycle length, phase sequencing, splits, offsets, and speeds that maximize bandwidth in both directions. PASSER was not designed to deal with oversaturated conditions. The link between queues at intersection approaches and the timing is indirect, the user has to specify the number of vehicles in the queue at the

beginning of each cycle and that number must be less than 6 vehicles. The model was designed for a single mode.

TRANSYT (Traffic Network Study Tool) is a signal timing tool developed by the Transportation and Road Research Laboratory in England in 1968. Version 7F was prepared specifically for the United States. It is a macroscopic, deterministic simulation and optimization model. It optimizes splits and offsets for a given cycle length. The model can use four performance indexes (PI) for optimization (FHA, 1991):

- disutility Index (DI) which is a linear combination of delays and stops;
- progression opportunities (PROS) defined as the number of green signals passed without stopping in one direction;
- PROS and then DI, and
- Ratio of PPROS and DI.

Until Release 8, the TRANSYT model was designed mainly for modeling unsaturated conditions. Release 8 introduced new capabilities to optimize for oversaturated conditions and four new optimization objective functions were introduced. More details are provided in Section 3.2.2.

A common characteristic of all the above models is that they were all originally developed for unsaturated conditions and for a single mode of transportation. Over the years, they have all been modified to handle oversaturated conditions, with limited results. However, although some of them have specific transit operation related parameters, they all remained single mode oriented.

2.4.2 Traffic Signal Control Systems

2.4.2.1 Pre-Timed (Time of Day Based) Control

Pre-timed control is the most popular signal control approach in North American cities. Signal timing plans are prepared by one of the models/tools described in the previous section for various times of the day. Normally there are at least three timing plans for each intersection, for the morning peak period, the afternoon peak period, and for off peak period. However, the number of timing plans is not limited, off peak periods could be broken down to periods during the day, evening, or late night/early morning, there could be separate plans for weekends, holidays, and special events. Timing plans are selected by a time clock and/or they could be overridden by an operator.

In comparison to adaptive control, beside its implementation cost, pre-timed control has two major advantages which are: transparency and technical simplicity, and substantial flexibility ability to implement changes to the control system.

Pre-timed control is relatively easy to comprehend and in technical terms it is relatively simple. These characteristics are often ignored but they are essential for cost effective set-up, operation, and maintenance. From the perspective of bimodal transportation management, however, the ability to make changes to pre-timed systems is the most important advantage. The transparency and simplicity of the control system allows operation modifications that are supportive of emerging transportation policy/planning shifts.

In terms of control effectiveness, pre-timed control may not be as effective as a properly tuned adaptive control. Pre-timed control is based on historical traffic counts and due to the likely changes in traffic patterns over time, implemented timing plans may not be the optimum plans unless they are updated frequently. Furthermore, they cannot adjust to unexpected and unplanned changes in traffic demand.

2.4.2.2 Adaptive Control

Adaptive signal timing and optimization models continuously monitor traffic operation, recalculate signal timing parameters, and adjust signal timing for existing traffic conditions. Although there are a number of adaptive traffic control system prototypes, only two systems, SCOOT and SCAT have been implemented on larger scale.

The first version of SCOOT (Split Cycle Offset Optimization Technique) was developed by the Transport Road Research Laboratory (TRRL) in the United Kingdom in 1981. The model continuously monitors traffic flows through a number of detectors placed along the approaches to the intersections and changes the green split, offset and cycle length to minimize stops and delays. The optimization function searches for signal settings at all intersections within the controlled network that will minimize the optimization function which includes delay and stops. Links (section connecting two intersections) can be weighted and the weights are used to bias the optimization process when increasing the priority of certain links is required.

SCOOT was originally designed to provide frequent and fast signal time changes as traffic conditions change. Therefore timing changes are made frequently, for example the stage duration can change in every cycle in increments from 1 to 4 sec, the offset can change in every cycle but by 4 sec and the cycle time is allowed to change with an increment of 4, 8, and 16 seconds normally every 5 minutes, or every 2.5 minutes if traffic condition changes very rapidly (SCOOT 1999).

SCATS (Sydney Coordinated Adaptive Traffic System) developed in Australia continuously monitors traffic flows at critical intersections and sets cycle length and splits to optimize operation. Offsets, compatible with the optimized timing plan are selected from a predetermined library of plans. The optimization function includes vehicle delay and stops.

The obvious advantage of traffic control, compared to pre-timed control, is its ability to adjust to unplanned traffic volume or travel pattern changes. However, despite this technical advantage, the implementation of adaptive control in the world has been relatively limited so far. There are three reasons for that: cost, complexity, and limited capability to deal with bimodal policy/planning objectives.

There are only a few adaptive systems that could be purchased and their implementation often requires significant signal control equipment modifications, extensive vehicle detection installation, and the purchase of new hardware and software. The magnitude of investment is in millions of dollars. Adaptive systems are relatively complex and the

initial set-up, operation, and maintenance of these systems require significant knowledge and experience. Unless traffic control agencies manage to develop in-house expertise, they will have to rely on continuous external support.

From the perspective of this thesis, however, the biggest disadvantage of existing adaptive systems is its limited ability to be adjusted to deal with multi-modal policy/planning objectives. This limitation has two aspects:

- **Difficult to get changes from the system provider:** Existing adaptive systems were developed with a single mode focus. They are relatively complex, and changing their focus to support emerging transportation policy/planning objectives is difficult. Shifting from the vehicle travel perspective to person travel perspective requires structural changes, which are unlikely. Changes are usually done as ad-on features which often have a limited value. For example, SCOOT has introduced the Transit Signal Priority function which increases the importance of transit vehicles in the optimization process thus results in less transit delay during unsaturated conditions. However, as the level of congestion increases, it gradually decreases priority to transit vehicles and eventually transit loses priority when it needs it the most from the policy/planning perspective.
- **Adaptive systems have a proprietary software:** The right for modifications are normally limited to the system provider.

2.5 Traffic and Signal Operation Simulation Tools

There are a number of simulation tools that can be used for the analysis of traffic/signal operation. However, the choice is very limited when the intention is to develop new signal control algorithms. VISSIM with its VisVap tool was used for simulation purposes in this thesis. VISSIM is a microscopic, driver behaviour-based multi-purpose traffic simulation program. The software was selected for the purpose of this research because it provides great flexibility in designing the roadway configuration, priority rules, volume distributions, and data collections. The VisVap signal control design tool allows the construction of new signal control algorithms, although its capability to do mathematical calculations is somewhat limited.

2.6 Summary and Conclusions

The term congestion is used as a generic definition of a traffic situation when demand is higher than capacity. Saturation and oversaturation are more refined definitions of congestion. In oversaturated condition the consequence of queue build up is more severe and it impacts upstream lane connections and intersections. In saturated condition the impact of congestion is limited to the bottleneck intersection.

The installation of traffic signals has almost a 100 years history. The original objective of signal control was to separate the travel of conflicting movements, thus it was primarily safety related. Most traffic signal control theories were developed with the objective to minimize vehicle delay with the assumption that the intersection has sufficient capacity.

From this perspective, queue management by signals has a somewhat different objective. The primary objective of queue management is to control the location of queue formations.

A number of traffic control models/tools are available for the analysis of traffic operation and for the calculation of signal timing plans. These models have two common characteristics: First, originally they were all developed for single mode operation, thus they all focus on vehicle delay and/or the number of stops. For example, minimizing person delay as an optimization objective has not been included in any of the models. Second, they have a limited applicability in oversaturated conditions.

In terms of signal control systems, two general approaches can be distinguished: the Pre-Timed (time off day) control approach and the Adaptive Control approach. Both control approaches have advantages and disadvantages, however, from the perspective of this thesis, the pre-time control has been selected because its operation is transparent, and it allows control algorithm modifications that could support emerging transportation policy/planning objectives.

3. LITERATURE REVIEW

3.1 Introduction

The section starts with a literature review of approaches to model and control traffic in congested conditions. A special focus is given to the review of pre-signal and gating logic development and application. Then the history of fuzzy logic and its applications to traffic control/management problems is investigated. Finally, literature dealing with the concept of traffic metering is reviewed.

3.2 Approaches to deal with Congested Conditions

3.2.1 Traffic and Signal Operation Analysis Models for Congested Conditions

Most traffic/signal operation analysis models were initially developed primarily for unsaturated conditions and they had very limited applicability in saturated or oversaturated conditions. Akcelik (1988) concluded that the 1985 HCM, the Canadian (Canadian Capacity Guide) and Australian (SIDRA) delay models provided reasonably similar delay values for unsaturated conditions, but they diverged significantly in oversaturated conditions.

A similar observation was made by Pitsiava et al. (1992) when they compared delay estimates of HCM (1985) and SIDRA. The authors concluded that in the case of

oversaturated conditions ($v/c > 1$), no specific rule or model can accurately estimate the values of delay.

Lawrence et al. (1989) compared the 1985 HCM method with two other models and concluded that the HCM delay compared reasonably well to other delay models such as Signal Operations Analysis Package (SOAP), Transyt-7F Release 5, but only for unsaturated conditions.

This is not surprising, since all of the compared delay formulas were based on steady state models but the oversaturated condition is not a steady state. The duration of an oversaturated condition is of great importance in the calculation of delay. Although all delay formulas include a time dependent component for oversaturated conditions, the duration of oversaturation was fixed in the 1985 HCM to 15 min, in SOAP to 60 min, while the Canadian and Australian models treated analysis time as an independent variable. Therefore, delay values for oversaturated conditions could not be the same.

Hurdle (1984) offered a critical analysis of delay models and described some of the limitations of these models by saying: “these models would seem to be useful as long as users do not expect much from them... if accuracy is desired, however, a new generation of models that take more account of variations in travel demand over time is needed. The use of such models would require more information about traffic patterns than users are accustomed to providing...”.

Engelbrecht et al. (1997) proposed a generalized delay model which was an improvement to the 1994 HCM delay model. Similarly to the Australian and the Canadian delay models, the proposed model allowed the use of any time duration for oversaturated conditions. The paper also points out the difficulties of verifying the formula in a real-world condition since collecting delay data from the field in oversaturated conditions is extremely difficult. Instead, the researchers used microsimulation to verify model results. Similarly to the HCM formula, the proposed generalized delay model still assumes that the demand distribution has a simple step function.

In year 2000, the new version of the HCM (2000) was published including the generalized delay model proposed by Engelbrecht et al. Although the new manual includes a procedure for analysing oversaturated conditions at isolated intersections, it does not provide a practical approach to signal timing for those conditions, or the analysis of non-isolated intersections. The HCM 2000 method's limitations for oversaturated conditions are explicitly pointed out: "The methodology does not take into account the potential impact of downstream congestion on intersection operation. Nor does the methodology detect and adjust for the impacts of turn-pocket overflows on through traffic and intersection operation." (HCM 2000, Pg.16-1)

In an overview of HCM 2000, Kittelson (2000) points out the need for future focus on oversaturated conditions: "...It is certain that evolutionary advancements and changes will continue to occur beyond the Year 2000. Future efforts will need to address more

complex problems such as the analysis of oversaturated facilities and networks.

Transportation system analysis techniques will need to be developed and refined...”

In conclusion, the literature review confirmed that the analytical models, developed for unsaturated conditions have only a very limited value in oversaturated conditions.

3.2.2 Approaches to Traffic Control/Management in Saturated and Oversaturated Conditions

In the early 90's the Transportation Research Board initiated a study dealing with oversaturated signalised intersections and in 1992 an NCHRP study was completed but not published (NCHRP, 1992). The study concluded that “no coherent approach to the problem of controlling an oversaturated traffic environment exists...” The study presented a new high level approach to signal design in the form of an Internal Metering Policy (IMP). However, the study did not develop a methodology to deal with oversaturated conditions; instead, it defined the objectives of a future IMP to manage the growth of queues to maximize the productivity of the roadway system. More specifically, the objectives were defined as:

- control queue formation to prevent spillback into intersections;
- fully utilize all available green time at the highest service rate;
- effectively use all existing roadway storage capacity;
- provide equitable service to competing traffic streams;
- minimize the number of vehicle stops; and

- minimize delay along the unsaturated elements of the network.

Two important conclusions can be made from the study. First, it is obvious that models describing the operation of isolated intersections in oversaturated conditions have a limited value. Second, traffic control in oversaturated conditions must focus on queue management. What the transportation profession needs are not only techniques to analyze oversaturated conditions, but new approaches to manage and control traffic under oversaturated conditions. In that spirit, a new research project titled Operation of Traffic Signal Systems in Oversaturated Conditions was formulated by Transportation Research Board (NCHRP 2007).

The objective of the proposed project is to research and develop a guide for operating coordinated traffic signals in oversaturated conditions on networks, including arterials. Some of the strategies and methods that should be addressed in the guide include metering, gating, turn restrictions, dynamic lane control and assignment, offset weighting, signal control priorities, and negative feedback progression (NCHRP 2007).

Although the development of traffic management tools for oversaturated conditions is long overdue, some of the concepts of dealing with oversaturated conditions are not new. Gazis et al. (1965) suggest that different algorithms should be used as demand increases from uncongested conditions to saturation and back to uncongested condition. At low demand levels (off-peak) the traditional minimum delay algorithm is appropriate. As the signal becomes saturated, a different algorithm developed to provide maximum capacity should be used. As demand starts to decrease, an algorithm that achieves the

simultaneous discharge of competing queues should be applied. They showed that at saturated intersections, the minimum delay is achieved by signal timing that cause competing queues to disappear at the same time. Finally, the minimum delay algorithm should be re-implemented.

Longley (1968) suggests that the objective of a control strategy for saturated conditions should include the following three elements:

- prevent or delay the formation of saturated condition;
- handle saturation when it happens; and
- recover from saturation as fast as possible.

When primary congestion occurs, the control strategy should be designed to avoid or to delay the formation of secondary congestion, and to minimize the impact of secondary congestion.

May et al. (1988) developed a fixed time signal plan for managing queues during oversaturated conditions. The method included the following elements:

- predicting the growth and decline of queues from downstream intersections;
- estimating the speed of the starting and stopping waves;
- identifying the most appropriate stage in which queues should arrive at the upstream intersection;
- identifying the most appropriate time in that stage during which queues should be present;

- adjusting the split between stages to allow for loss of throughput during blocked stages; and
- adjusting the cycle time as necessary.

May (1991) discusses a range of queue management techniques including options for phases to be blocked, timing of the arrival queue in the blocked phase and the adjustment of split and cycle times. The queue management strategy is based on the assumption that the speed at which the front of the queue (end of the queue) moves backwards can be predicted with some degree of accuracy. The speed of the starting and stopping waves can be determined by observation and by applying the principles of traffic flow theory. However, more than one formula can be found in the literature for the speed of the starting wave and all provide significantly different results.

Quinn (1992) writes that in congested operations both capacity and operational efficiency are degraded and strategies are needed to manage vehicular queues. During unsaturated conditions, the objective of signal control is to facilitate safe movement and minimize delay. However, during saturated conditions, the objective of signal control has to be the maintenance of maximum throughput. Conventional procedures for optimizing fixed-time signal control, including TRANSYT 7 F (prior to Release 8) deteriorate rapidly when severe congestion persists and alternative strategies are required to manage queues. The author suggests a few solutions to the problem, the first being the manual override of signal operation by traffic police as is the standard practice in the world's most congested cities. A second solution is to provide 'reverse' offset for signal coordination. The

conventional offset used for unsaturated arterials allows forward progressions of platoons of vehicles, the start of the green time of a downstream intersection is determined by the start of the green time of the upstream intersection. With reverse offset, the start of the green phase at an upstream intersection is determined by the start of the green at a downstream intersection. The objective of reverse offset is to minimize queue spillbacks.

TRANSYT developers were aware of these problems and to facilitate the analysis and control of oversaturated conditions, TRANSYT 7F completely restructures the simulation process in Release 8 in 1998. Previous versions of the model were based on a linkwise approach where simulation begins at a specific link, completes all simulation steps, and then proceeds to the next link sequentially until all links in the network are simulated. Queues were described as vertical queues. This approach did not allow the effects of spillback occurring over time to pass from one link to another (Li et al., 1999). Release 8 implemented a stepwise approach where the complete network is simulated at each time step. Vehicular queues were described as horizontal queues. This allowed the transfer of spillbacks from one link to another over time.

Li et al. (1999) tested TRANSYT 7F Version 8 on an oversaturated diamond interchange and concluded that the all new objective functions for oversaturated conditions produced better timing plans compared to the objective functions of previous releases. The new objective functions significantly reduced the duration of spillback conditions and generated lower delays.

Gal-Tzur et al., (1993) developed a method for designing signal programs for congested urban networks based on limiting the volumes entering the network. The procedure was based on selecting links that served as a buffer thus the strategy was called queue location management (QLM). A mixed-integer linear programming (MILP) model was developed for calculating the durations of green at all intersections and it was integrated into TRANSYT.

Rouphail et al. (1996) published an interesting paper in which they questioned the traditional traffic engineering wisdom by asking the following question: “More signal capacity with less green?” The authors analyzed the operation of closely spaced pair of intersections with heavy traffic flow and suggested that by limiting the green time at the upstream higher capacity intersection feeding the downstream bottleneck intersection the capacity of the intersection pair can be increased. Without metering traffic at the upstream intersection, queue interactions between the two intersections may reduce the saturation flow rate at the upstream intersection and consequently the discharge rate at the bottleneck intersection. The two-intersection system set up is shown in Figure 3.1.

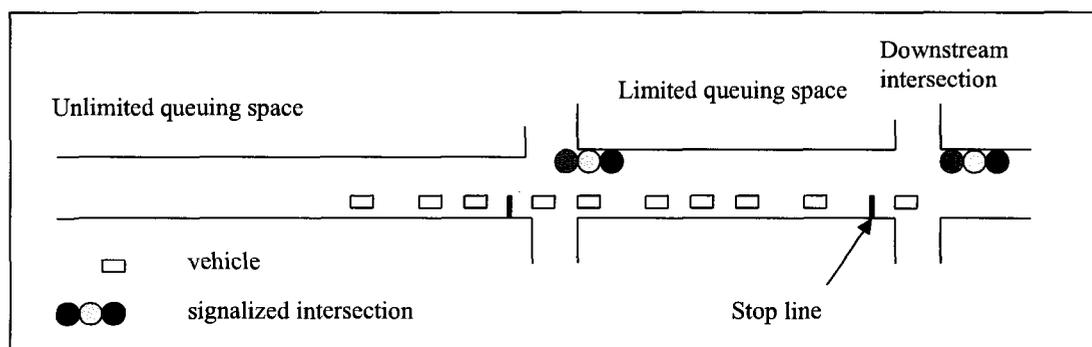


Figure 3.1: Two-Intersection System

Queue interaction will happen in two circumstances. In the first circumstance the upstream saturation flow rate is reduced when the lead vehicle in the platoon encounters a queue that requires the platoon to slow down and discharge at a lower rate. In the second circumstance the upstream saturation flow rate is reduced when the upstream signal discharges more vehicles than the storage capacity between two intersections.

The authors developed an analytical procedure to estimate the capacity reduction of the un-metered traffic control and to calculate the green time at the upstream metering intersection that eliminates queue interactions. The procedure is an iterative one because the length of the downstream queue depends on the upstream saturation discharge rate, and the upstream saturation discharge rate depends on the length of the downstream queue. The main steps of the procedure are shown in Figure 3.2.

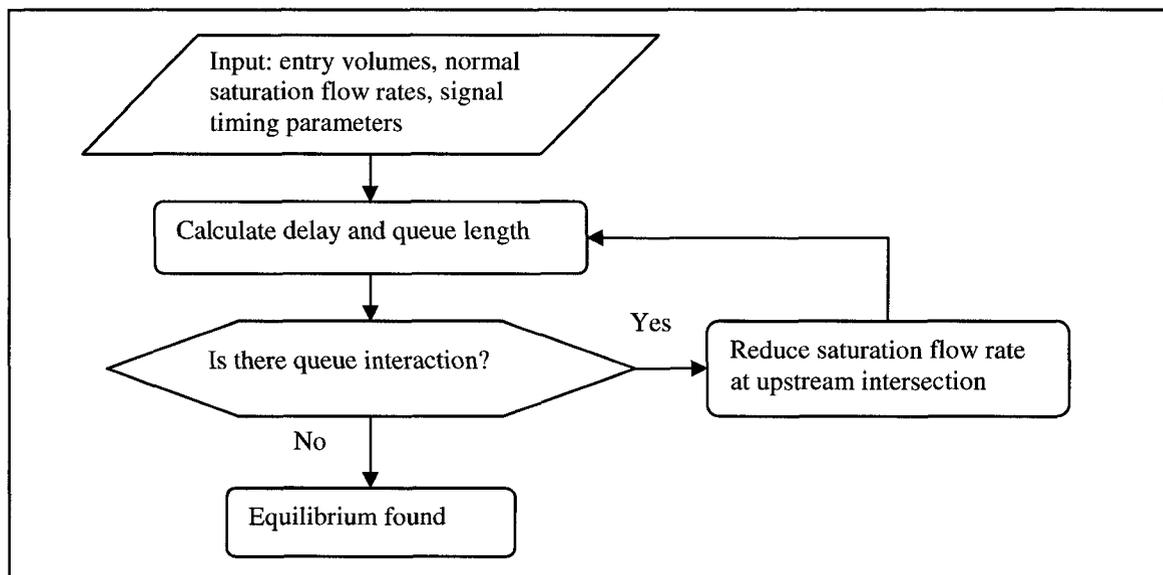


Figure 3.2: Simplified Analysis and Signal Timing Procedure by Rouphail et al. (1996)

The delay and queue length calculation included both the deterministic and random components and the authors concluded that queue interactions could occur even when the upstream intersection is slightly under capacity. By using an example in the paper, the authors showed that the queue interaction resulted in a 10% capacity loss at the downstream intersection. By reducing the green times at the upstream intersection, the queue interaction was eliminated and maximum capacity at the downstream intersection was restored. Total delay (at both intersections) varied with the offset between the two intersections; however, it was always lower without queue interactions. The authors concluded that the use of simultaneous offset settings and metering policies at the upstream signal were effective in terms of maximizing capacity and minimizing delay.

Abu-Lebdeh et al. (1997) used the genetic algorithm approach to develop a control mechanism for oversaturated arterials. The objective was to maximize throughput and the control principles were the following:

- manage queue formation and dissipation so that no green time is wasted i.e. traffic is discharging with the maximum discharge rate;
- manage queue lengths so that no upstream intersection blockage occurs; and
- account for the non-steady state characteristics of the system.

The authors provided a mathematical formulation of the above three principles in the form of a dynamic discrete-event model. However, instead of using techniques such as dynamic programming, they used genetic algorithms to calculate signal parameters.

Simulation results on a simple one directional arterial with five signals and no turns suggested that the method logically changed (increased) offsets as the queue length at the intersection approach increased. Negative offsets were applied properly and it also reduced green times at upstream intersections as queue lengths increased. Although the developed principles could be applied to any number of intersections, computer memory and CPU time still represented a significant limitation.

In a follow up paper, Abu-Lebdeh (2000) expands the above method to two way arterials that are oversaturated in both directions. However, the presented genetic algorithm is structured to find optimal control in one direction only, the primary direction. Progression is provided only in the primary direction, in the secondary direction the method only manages queues so that queue spillbacks are minimized. The procedure combines mathematical optimization with engineering judgment. The formulation of two-way control and queue management presented in the paper is far more complex than the one-way control.

Clegg et al. (1999) suggest that Urban Traffic Control, more specifically signal timing could be used to achieve Traffic Demand Management objectives and not only the traditional goals of maximizing throughput and minimizing travel times. In three European cities fixed time signal timings have been implemented with various objectives. In York, UK the objective was to improve transit travel times along a corridor by using the gating technique. In Porto, Italy the objective was to reduce overall congestion, improve transit travel times along selected routes and discourage cars entering pedestrian

sensitive areas. In Thessaloniki, Greece the objective was to reduce vehicular congestion and improve transit travel times along selected bus routes. It is important to note that congestion reduction was not done by the traditional approach of trying to maximize capacity. A new technique called Management of traffic USIng Control (MUSIC) was developed to influence trip assignments along the network.

Simulation results suggested that considerable gains can be achieved by rerouting traffic by fixed time signal operation designed to influence route choices. The authors argued that these gains cannot easily be achieved by using traditional responsive signal setting-policies and fixed signal timing policies may be more effective when networks are highly congested. The study concluded that the MUSIC approach showed the possibilities of achieving TDM objectives by developing special signal timing plans, however the MUSIC approach proved to be extremely time consuming and thus impractical. The need for future research in this area was highlighted.

Lieberman et al. (2000) formulated a dynamic traffic control model RT/IMPOST (real-time/internal metering policy to optimize signal timing) designed for oversaturated arterials. The signal phase duration was set to restrict traffic volumes at intersections servicing oversaturated approaches with the objective to control and stabilize queue lengths. Signals were coordinated to fully utilise the available storage capacity. The control had the following three objectives:

- maximize throughput by avoiding queue blockages, intersection starvation, and by managing queue formation to maximize the discharge rate;

- maximize storage capacity to limit the area where queues were formed, but avoid queue blockages; and
- provided equitable service to allow adequate service to all approaches without generating disproportional delays.

A mixed-integer linear programming MILP tableau has been formulated to provide optimal values for signal offsets and queue lengths for each approach. A nonlinear (quadratic) programming formulation has been used to calculate arterial green time durations of each signal cycle so that the actual arterial queue lengths on each saturated approach would continually approximate the optimal queue lengths determined by the MILP formulation.

Timing plans calculated by the RT/IMPOST, Synchro, Passer II, and TRANSYT were compared by using WATSIM on an arterial with six signalised intersections. Traffic volumes varied over time, the simulation started and ended with unsaturated conditions and had an oversaturated peak. Except for the first unsaturated timing plan, the RT/IMPOST significantly outperformed the other methods.

During the last decade, a number of papers were published on the application of genetic algorithms for traffic control and queue management in oversaturated conditions.

Abu-Lebdeh et al. (1997) formulated the problem of control as a throughput maximization problem subject to state and control variables and used a genetic algorithm to search for the best solution. The algorithm was applied to an arterial with five

signalized intersections. Abu-Lebdeh et al. (2000) further developed the concept and applied it a two-way saturated arterial.

Girianna et al. (2004) used generic algorithms to design signal coordination for oversaturated networks. The algorithm provided signal timing that was responsive to fluctuations in traffic volumes. One of the major issues pointed out was computational time. The authors also listed a number of areas for further research.

The literature review confirmed that the control of oversaturated conditions is significantly more complex than the control of unsaturated conditions. It also revealed that the focus of control must be on queue management and not on the minimization of intersection delay. It also seems that the research changes focus from analytical models to methods of artificial intelligence, including genetic algorithms. Furthermore, the control must be dynamic and therefore microscopic simulation tends to be a more feasible method of evaluation than the reliance on analytical models.

3.2.3 Pre-Signals and the Gating Logic

In the process of dealing with oversaturated conditions, the need to manage traffic volumes and the formation of queues has been clearly established. A specific technique used for multi-modal queue management is the gating logic. The objective of the gating logic is to relocate traffic queues from sensitive areas to links that are designated for queuing.

One of the first implementations of the gating logic, dubbed “sluicing”, were in Düsseldorf, Germany in the early 70’s (NATO, 1976). Unfortunately no specifics were found.

In the UK, the concept of queue management by using pre-signals was suggested the first time in 1991 (UK DOT, 1997). The first implementation at Park View Road on Uxbridge Road/High Street happened in 1994. The scheme had the following bimodal objectives (Oakes et al., 1995):

- enable buses to overtake the queue at the two lane roadway and rejoin the general traffic stream on the uncongested final approach to the downstream signal;
- provide a flow metering strategy that will respond to the level of congestion and keep the queue to the downstream bottleneck intersection to a minimum and store the residual queue on the two-lane roadway upstream from the pre-signal; and
- the overall system should not affect the capacity of the downstream bottleneck intersection.

It was recognized that the control system required to meet the design objectives would require a completely new approach to signal controller operation. The queue relocation algorithm had to be dynamic and constantly adjust the green time based on conditions measured at both upstream and downstream nodes. The downstream signal (main signal) operation remained unchanged and it was optimized for that location. The upstream signal (the pre-signal) had to monitor the operation of the downstream controller operation and make adjustments based on the queue formation between the two signals.

The pre-signal was coordinated with the main signal through offset timing, and the pre-signal green time was calculated by a special algorithm developed for this purpose.

The duration of the green phase at the pre-signal was calculated based on the difference between vehicle counts at the pre-signal and main signal. The system was regularly recalibrated by using queue detectors. The recalibration was needed because the distance between the pre-signal and the main signal was relatively long (approximately three blocks), and the traffic sources and sinks within this distance had to be accounted for. The pre-signal control also required a damping mechanism to prevent the generation of shock waves due to the highly variable downstream signal operation.

Oakes et al. (1995) provide a relatively detailed description of the relevant factors at the pre-signal at Park View. The schematic site layout is presented in Figure 3.3.

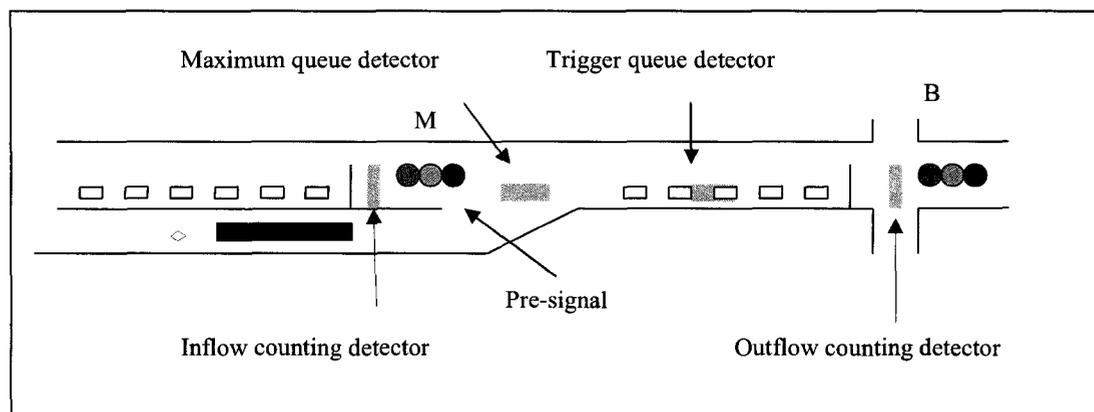


Figure 3.3: Elements of the “Park View” Pre-Signal Set-Up

The main elements of the pre-signal physical set-up are the following:

- inflow counting detector counts the number of vehicles (N_2) crossing the stop line at the metering signal ‘M’ during one cycle;

- outflow counting detector counts the number of vehicles (N1) crossing the stop line at the bottleneck signal 'M' during one cycle;
- trigger queue detector detects the presence of a queue. At the Park View application this has been done by three detectors. Actuation of any two of the three detectors for a predefined period triggered the queue relocation mode; and
- maximum queue detector detects the presence of too long queues.

The operational algorithm, based on the description in the paper, is further discussed in Section 8.4.1.

In SCOOT Version 2.4 a number of features were introduced to tackle oversaturated conditions (SCOOT, 2007). The most important one is the 'gating' logic. Within the SCOOT gating logic, links can be identified as bottleneck links and gated links. The bottleneck links represent the lowest capacity locations in the network. The gated links are used to store the queues which would otherwise block the bottleneck link. Traffic volumes are controlled at the gating links to control volume into the bottleneck link. For the bottleneck link, the traffic manager/operator defines the critical degree of saturation through the bottleneck which is used to trigger the gating action. The gating effect is implemented by changing the splits at the gating signal. Further details on the gating logic are not available.

In SCATS, timing plans providing a 'gating' function can be also selected (Al-Mudhaffar, 2006) however no further details on its operation or its effectiveness was found.

Hounsell et al. (2005) described a number of queue relocation techniques by using pre-signals. The focus of the paper is on developing an analytical approach to evaluate the effectiveness of various types of pre-signal setups. The paper concludes that by using queue relocation, it is possible to provide significant transit benefits without having negative impact on general flow.

The description of the applied control algorithm is very cursory. Instead, the authors suggest that “special attention should be paid when implementing the signal control logic so that both pre-signal and main signal operate efficiently. The evaluation was done by using the deterministic queuing theory and it was developed for fixed time control and unsaturated conditions.

Kumara et al. (2007) used microscopic simulation to evaluate a bus priority system based on queue relocation with pre-signals. The paper points out the following four benefits of queue relocation with pre-signals:

- time savings to buses by allowing the bus to approach the bottleneck intersection without delay;
- additional time savings at the intersection downstream from the pre-signal;
- minimizing conflicts for left turning (right turning in UK) buses at the intersection downstream from the pre-signal; and
- congestion free movement for buses and general traffic downstream of the pre-signal.

The evaluation was carried out at two locations (Hanwell Bridge and Edgware Road). Both locations were controlled by SCOOT. The authors suggest that the complexity of the control depends on the actual purpose of the pre-signal and on the detector placements.

The provided controller algorithms reveal only the high level general approach. The pre-signal is green until an approaching bus is detected. After detecting the bus at an upstream detector, the pre-signal turns red for general traffic for a specific duration. After the pre-set time elapsed, the pre-signal is turned back to green. The pre-signal responds to bus actuation immediately, thus there is no coordination between the pre-signal and the downstream main signal.

The application at the Hanwell Bridge has two objectives, to fully utilize the green time at the main signal and also to provide bus priority. The control logic is more complex than the logic at Edgware Road. The queue relocation algorithm is turned on when the level of saturation between the pre-signal and the main signal reaches 94% for three consecutive cycles. In contrast to the Edgware Road application, the operation of the pre-signal is coordinated with the main signal.

The paper suggests three alternative approaches to determine the duration of the pre-signal green time, as a function of entering and leaving traffic, as a fixed value, or as the function of the main approach green time. There are no specifics about the method used during the evaluation.

The City of Ottawa has implemented queue management with the gating logic at a number of intersections. However, due to the unavailability of practical queue management signal control techniques that could be built into traditional signal control systems, the effectiveness of these measures is currently limited (Mucsi, 2005).

In conclusion, some information on queue management schemes with the gating logic and pre-signals implemented in the UK was found but details about the control logic are lacking.

3.3 Fuzzy Logic in Traffic Control

The application of fuzzy logic in the Western world is relatively new. The Western philosophy, engineering training, thinking, and consequently engineering applications are based on foundations that were laid down by ancient Greek philosophers. But is that the only and best possible foundation for engineering thinking?

The following literature review starts with a brief background of the philosophical and historical aspects of fuzzy logic and ends with the description of a few fuzzy logic applications in traffic engineering.

3.3.1 Philosophical and Historical Aspects of Fuzzy Logic

Modern mathematics and logic is based on the bivalent or binary logic. Bivalent logic states that every proposition must be either true or false, without a third alternative

(Brule, 1984). In mathematical terms everything is either A or not A, in the field of electronics it is 1 or 0, in social science it is 'good' or 'bad'.

The history of the bivalent logic goes back to the ancient Greek philosophers. Parmenides (ca. 510 B.C- ca. 480 BC) suggested that there is nothing that 'can be' and that 'cannot be' at the same time by describing the bivalent principle of non contradiction: "that which is *is* and that which is not *is not*" (Campbell, 2002). Similarly, the Aristotelian logic looks at the world in a bivalent manner. Aristotle (384 BC – 322 BC) suggested that "...it will not be possible to be and not to be the same thing" (Aristotle, 1952).

Although the bivalent logic became the foundation of science in the Western world, its acceptance has not been universal. Heraclitus (ca. 535 BC – ca. 475 BC) questioned the purely bivalent logic by asserting that "I am as I am not" (Lavine, 1984). Furthermore, Plato (ca. 428BC-ca. 348 BC) suggested that there is a "third region", beyond true and false (Lavine, 1984). More recent philosophers such as Hegel (1770-1831), Marx (1818-1883), and Engels (1820-1895) followed this thinking and in the early 1900's Lukasiewicz, a Polish mathematician, described a three valued logic with the supporting mathematics (Lejewski, 1967). The third term was described as "possible" and he assigned a numeric value to it.

The influence of the bivalent logic on the Eastern philosophy has been much less prominent. The philosophy of Buddhism (500 BC) was based on the view that the world is filled with contradictions and almost everything includes some of its opposite.

Contrary to Aristotle's bivalent logic, in Buddhism things can be A and not A at the same time. This grey or fuzzy view in Eastern belief goes from the Chinese Taoism to the modern Zen in Japan (Kosko, 1993).

Fuzzy logic as known today was defined by Lotfi Zadeh in his seminal work *Fuzzy Sets* (Zadeh, 1965). A fuzzy set is a class of objects with a continuum of grades of membership. Such a set is characterized by a membership (characteristic) function which assigns to each object a grade of membership ranging between zero and one. Although there are many similarities between the classical set theory and fuzzy set theory, fuzzy set theory is not based on the law of excluded middle and the law of contradiction (Yen, 1999).

Fuzzy logic is an infinite-valued logic, and it is based on the description of falseness and truth by using real values between 0.0 and 1.0. Zadeh described the mathematics of fuzzy set theory and by extension fuzzy logic. He showed that the fuzzy definition of truth values allows the development of effective algorithms that are based on vague logical statements and vague data (Teodorović, 1998).

Zadeh applied the fuzzy set theory to control problems and proposed a control logic for an air conditioner which was based on linguistic control commands such as "work a little faster when it gets hotter or slower when it gets colder", instead of developing a mathematical model for controlling the operation (Kosko, 1993).

By analysing the operation and control of complex systems, Zadeh formulated the “principle of incompatibility” which states that “ ...as the complexity of a system increases, our ability to make precise and significant statements about its behaviour diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost mutually exclusive characteristics...” (Kosko, 1993). Zadeh suggested that in those circumstances, fuzzy logic based control has the potential to outperform mathematical based controls (Kosko, 1993).

One of the first successful applications of fuzzy logic has been in a cement kiln in Denmark in 1977 (Jamshidi, 1993). While scientists in the Western world tended to stay away from the fuzzy logic approach, engineers in Japan enthusiastically applied it in many commercial applications. The first subway system utilizing a fuzzy logic-based automatic train operation control system was built in Japan in 1987. Today, fuzzy logic is used in washing machines, air conditioners, anti-lock brakes, auto transmissions, cameras, elevators, industrial processes, and in many other systems (Kosko, 1993).

3.3.2 Fuzzy Logic in Traffic Control

The first documented attempt to apply fuzzy logic in signal control was carried out by Pappis et al. (1977). The researchers developed a fuzzy logic control algorithm for a single intersection of two one-way streets. Besides being able to provide better control with the fuzzy logic controller than with the alternative vehicle-actuated controller, the

authors pointed out the convenience of developing a control algorithm based on linguistic terms.

Beauchamp et al. (1997) developed two control algorithms for an isolated unsaturated intersection that used fuzzy logic (FL). In the first algorithm called Fuzzy Logic Controller for Traffic Systems (FLC-TS), fuzzy logic was used instead of the traditional gap based logic to determine whether or not to terminate a particular phase. In the second algorithm called Phase Sequencer plus FLC-TS (PS+FLC-TS), the authors introduced an additional function based on fuzzy logic, the selection of the next phase to be serviced. The control framework is presented in Figure 3.4.

The two fuzzy logic based algorithms were compared with the traditional fixed time controller logic by using a proprietary microscopic simulation program. The testing was very limited in terms of both duration and in terms of the input values. The simulation was carried out for only 240 seconds due to the limitations of the microscopic simulation software, and only two volume levels were used. Based on this limited test, the mean delay for the two fuzzy logic controllers was less than for the fixed timed controller. The mean number of vehicles that cleared the intersection was higher for the two fuzzy logic controllers. Interestingly, there was very little or no difference between the two FLC-TS and the PS+FLC-TS controllers.

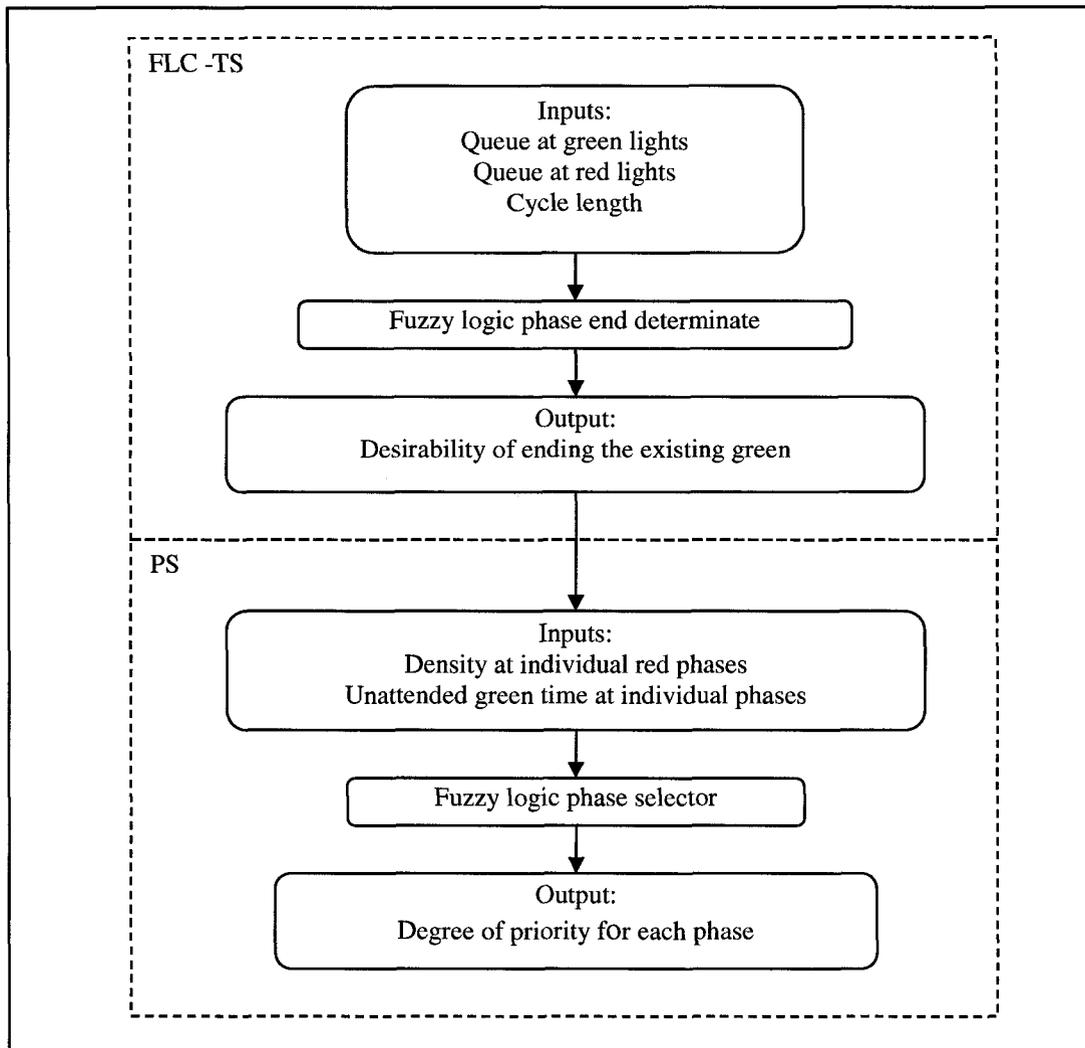


Figure 3.4: The Framework of Beauchamp et al. (1997) Fuzzy Logic Controller

Niittymäki and Kikuchi (1998) demonstrated the application and multi-objective capability of fuzzy logic for the control of a pedestrian crossing signal. The control had three objectives, to minimize pedestrian wait time, to minimize vehicle delay, and to maximize safety to vehicles and pedestrians. The pedestrian delay and safety objectives were based on the logic that pedestrian green should be displayed as soon as possible when a pedestrian is detected but not necessarily immediately. The objective of

minimizing vehicle delay was based on the logic that the pedestrian green time should be as short as possible. The vehicle safety objective was based on the logic that the termination of vehicular green while a platoon of vehicles is approaching should be minimized. It is obvious that some of these objectives are contradictory and one of the advantages of FL based control is in its ability to handle trade-offs.

The Fuzzy Logic controller operation was compared to two types of conventional demand-actuated control, the Conventional controller operation and Modified controller operation through micro-simulation. The difference between the Conventional and the Modified control was in the maximum length of extensions. The Conventional operation allowed 4 second extensions up to a maximum phase length of 35 seconds. The Modified control allowed only one extension of 4 seconds. The Modified control represented an attempt of using traditional controller functions to provide pedestrian “friendly” operation.

Both the Conventional and Modified controller operation had pedestrian detectors and extension loops for vehicles. The signal rested in vehicular green until a pedestrian activated the pedestrian detector. Both vehicle and pedestrian phases had minimum green times and the vehicle phase could get an extension by a passing vehicle.

The comparison of the three control methods was done for three levels of pedestrian activity. The paper concludes that the Fuzzy Logic controller generally provides

pedestrian friendly control while achieving lower vehicle delays than either the Conventional or the Modified control.

According to Niittymäki (2001) one of the advantages of fuzzy logic in traffic signal control is the ability of fuzzy logic to support various objectives of traffic control. While maximizing vehicular capacity or minimizing delay and/or stops is the main objective of traditional traffic signal control algorithms, the objectives of fuzzy logic controllers are very flexible. This flexibility supports various transportation policy objectives, such as pedestrian friendly signal control, separate signals for cyclists, transit signal priority, heavy vehicle priority, environmental sensitivity, etc.

The paper describes a simple scenario with three control objectives: maximize safety, minimize delays, and minimize impact on the environment. Safety and the impact on the environment can be maximized by minimizing stops, thus increasing the cycle length. However, delay can be minimized by reducing the cycle length. With traditional signal control, it is difficult to handle all three objectives because it is often too complex to describe the relationship between the objectives. On the other hand, the FL approach allows multi objective signal control without requiring the detailed mathematical formulas describing the connections between the objectives.

The paper continues with the conceptual level descriptions of fuzzy logic control for isolated pedestrian crossings, two-phase vehicle control, multi-phase vehicle control, isolated traffic control at arterial intersections, and transit priority control.

In a follow up paper, Niittymäki et al. (2001) discuss the development and test results of various fuzzy logic control algorithms on an unsaturated and isolated arterial intersection. The control objective included the minimization of mean delay, the maximization of safety and the efficiency of the flow along the arterial.

Two fuzzy logic control strategies, Basic FL, and Enhanced FL were developed and compared with the traditional vehicle actuated (VA) control and with LHOVRA which represents the typical Swedish signal control method. The Basic FL control determined only the end of the green phase. For inputs it used the number of arrivals on the green phase and the number of vehicles in queue behind the red light. The output variable specified the green phase ending moment.

In addition to the phase termination function, the Enhanced FL control strategy had three more functions. The objective of the first additional function was to determine the sequence of phases. The objective of the second and third additional function was to minimize the chances of phase termination when vehicles were in the dilemma zone. The second additional function applied for heavy vehicles (detection zone up to 300 meters from the stop line), and the third function applied to passenger cars (detection zone between 50 and 120 meters from the stop line).

The control functions were organized to operate at two levels. The first level includes the next phase function and the phase termination function. The second level includes the phase function. For example, after the green termination rules have decided to terminate

the phase (from the delay perspective), the second and third level determines the exact moment when to do it from the safety perspective.

In terms of the mean delay along the major road, the Enhanced FL control outperformed the LHOVRA control at high volume levels. At lower volume levels, the LHOVRA provided the least delay. The Enhanced FL controller outperformed the basic FL control at all levels, however, the difference between the two diminished as the degree of saturation increased. At high levels of saturation, the difference between the Basic and Enhanced FL control diminished. In terms of the side street delay, the Enhanced FL control consistently outperformed the other two, while the Basic FL control consistently outperformed the LHOVRA control. The fuzzy phase selector improved the performance of the fuzzy control at low and medium demand. The safety functions' effect on the delay performance of the Enhanced FL control was not significant. The authors concluded that FL control was able to provide an efficient control while dealing effectively with trade-offs.

The objective of the research of Zhang et al. (2005) was to test FL control for oversaturated conditions. This has been done on an isolated intersection, with all 8 NEMA movements/phases. Similarly to previous research, fuzzy logic was used to determine whether or not to terminate the existing green phase. The interesting new element in this research was the application of fuzzy logic control to oversaturated conditions. Five volume levels were used, from 20 % to 100% where the upper limit represented substantially oversaturated conditions.

The fuzzy logic control (FL) was compared to pre-timed control and fully actuated control. In the comparison of the three control approaches five MoP were used, network delay, network speed, % stops, network time in queue, and network throughput to demand ratio. In terms of network delay, network speed, network time in queue, the FL outperformed the other two methods for volume levels of 60%, 80%, and 100%. In terms of % of stops, FL control performed the best at volume levels of 20% and 40%. However, actuated control performed the best at volume levels of 80% and 100%. In terms of the throughput to demand ratio, the FL control outperformed the other two methods.

In conclusion, there are a number of examples of using fuzzy logic for signal control purposes. In most cases, fuzzy logic was developed to replace the existing controller logic. Most studies concluded that, besides providing an efficient signal control, fuzzy logic also allows an easier handling of trade-offs in multiobjective control.

3.4 Metering

Ramp metering applies to on-ramps to freeways but the concept of metering may be applicable to signalized arterials and networks as well. Since the objective of this research is to investigate the metering concept for arterials, it is useful to review the fundamental elements and experiences of freeway ramp metering.

Ramp metering can be defined as an approach to manage traffic entering a freeway by using traffic signals with the objective to optimize traffic operation. The primary objective of optimization is to maximize traffic flow and minimize total travel time on the

freeway, secondary objectives may include crash reduction, limiting the queue length or delay at the signalized ramps, or the equitable treatment of motorists.

Ramp metering has a 50+ year history in North America. Implementation started in the late 1950's in Los Angeles, Detroit, and Chicago and based on a survey carried out by The Urban Transportation Monitor (2003), there were at least 21 North American urban areas with functional ramp metering systems in 2003. In addition to on-ramp metering, freeway-to-freeway connector ramp meters were also implemented in several areas including Minneapolis, San Antonio, and San Diego (Pearson et al. 2003).

Ramp metering can be based on historical, real-time, and predicted traffic volume data. In terms of strategies, ramp metering algorithms can be fixed time, local traffic-responsive, and system-wide traffic responsive (Pearson et al., 2003).

There are a number of ramp metering algorithms, however, only three strategies will be discussed here including the two most popular approaches, the demand-capacity and ALINEA methods, and the Fuzzy Logic based algorithm.

Papageorgiou et al. (1997) provided a concise description of the two most popular ramp metering methods in North America. The fundamental variables used in ramp metering are depicted in Figure 3.5 and they include q_{in} (the volume before the merging area), q_{cap} (the capacity after the merging area, the corresponding occupancies o_{out} , and r (metered volume on the ramp).

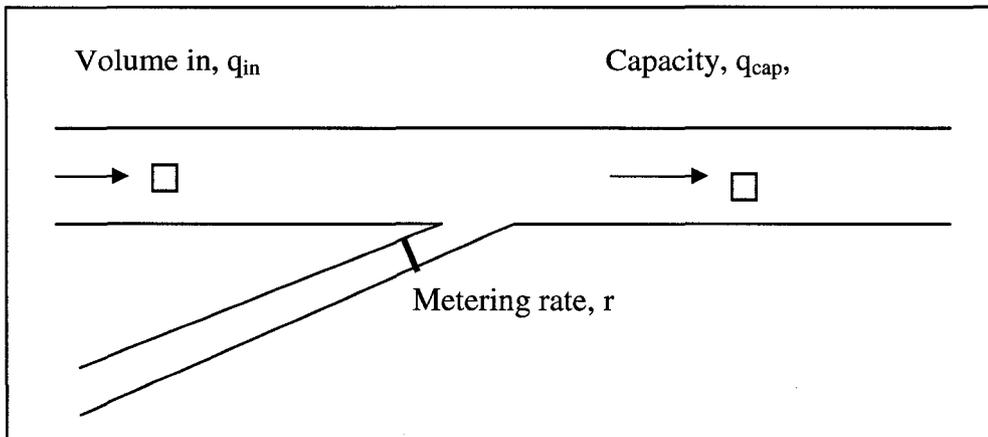


Figure 3.5: Elements of Ramp Metering

With the demand-capacity strategy, the ramp volume r is determined by calculating the difference between the downstream capacity q_{cap} and the upstream volume q_{in} . However, since the volume measurement q_{in} is not sufficient to describe congested conditions, occupancy data o_{out} must be also collected. The ramp metering algorithm is shown in 3.1.

$r(k)$ – ramp volume in time step k (veh/hr)

$$r(k) = \begin{cases} q_{cap} - q_{in}(k) & \text{if } o_{out} \leq o_{thres} \\ r_{min} & \text{otherwise} \end{cases} \quad 3.1$$

q_{cap} – capacity or desired flow (veh/hr)

q_{in} – measured flow at the upstream location

o_{out} – occupancy at the downstream location

o_{thres} – occupancy threshold at the downstream location

r_{min} – minimum metering volume for capacity conditions

The ALINEA (Asservissement LINEaire d'entrée Autoutiere) method calculates the metered volume based on the metered volume from the previous cycle and the "reserve

capacity. The “reserve capacity” is estimated by comparing the occupancy measurement at the downstream intersection with a threshold occupancy value. The ALINEA formula is presented in 3.2.

$$r(k) = r(k-1) + K_R \times [o' - o_{out}(k)] \quad 3.2$$

$r(k)$ – ramp volume at time step k

$r(k-1)$ – measured (not estimated) ramp volume at the previous time step

K_R – regulator parameter

o' – set value for occupancy

$o_{out}(k)$ – occupancy at the downstream location at time step k

Taylor et al. (1998) experimented with the fuzzy logic based approach and tested the design at multiple ramps along a freeway. The objective of metering was to minimize total travel time while maintaining acceptable ramp queues. The algorithm included seven inputs including occupancy measurements at five locations and speed measurements at two locations. Five fuzzy membership functions were used for each input parameter including: very small (VS), small (S), medium (M), big (B), very big (VB). The rule base was developed based on human experience and contains 17 rules. Separate weights were used for individual rules to increase or decrease the importance of the rules during the tuning process. The output of the fuzzy logic reasoning was the metering rate with five fuzzy classes, very small (VS), small (S), medium (M), big (B), and very big (B).

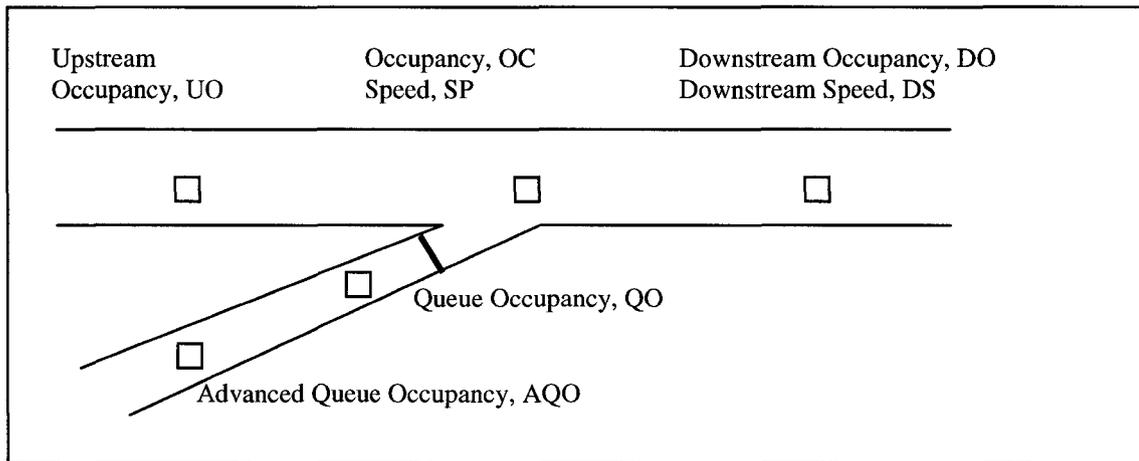


Figure 3.6: Input Data Collection Locations

The testing was carried out in a FRESIM simulation environment, and the fuzzy logic approach was compared to three ramp metering methods, the clocked metering, demand-capacity metering, and the speed metering.

The performance criteria included total system travel time while observing the constraints of maximum queue lengths on the ramp. Results showed that as demand exceeded capacity, the advantage of the fuzzy controller increased in terms of balancing between freeway operation efficiency and ramp queues. However, the authors also pointed out that the superior performance of the fuzzy logic controller could not be attributed solely to the nature of fuzzy logic, comprehensive inputs such as detailed ramp occupancy data also contributed to the better results.

Pavis et al. (1994) describe the operation and the benefits of metering traffic at the Lincoln Tunnel. A fixed metering rate was applied to one car and one transit lane which

merged after the metering point. The metering control resulted in a smoother traffic flow through the tunnel and an increase of capacity by 2000 passengers per hour.

It can be concluded, that the metering concept is not new, but its implementation has been limited to uninterrupted facilities.

3.5 Summary and Conclusions

Currently there are no widely accepted mathematical models that describe the operation of oversaturated intersections or networks. Traffic analysis models developed for unsaturated conditions focus on delays and stops and they are ineffective for oversaturated conditions. The enhancement potential of these models to describe intersection operation in oversaturated conditions is very limited. Models for unsaturated conditions are steady state models, but congestion cannot be described by constant demand. Queues develop, grow, reach a maximum, and then decline. Queue blockages have a profound impact on operations in oversaturated conditions, but the comprehensive inclusion of queue blockages in analytical models is quite challenging.

Traffic management for oversaturated conditions has been researched and documented by a number of researchers. It can be noticed that there is much less focus on the development of analytical models, instead the focus has shifted to methods that are at least partly based on heuristics.

The control of oversaturated conditions is complex by itself, and the efforts to develop bimodal control techniques have been limited. The literature search identified only a few

papers that describe and evaluate queue management techniques designed to improve the movement of transit vehicles. However, details of control algorithms and its effectiveness are lacking.

The literature search identified a number of papers describing and/or evaluating control algorithms based on fuzzy logic. Fuzzy logic has an inherent capability to deal with trade-offs. At intersections, fuzzy logic control can be used to balance the need to increase the lengths of a green signal and the need to terminate the green signal based on the queues and/or delays at the red signals. Similarly, at ramp metering, fuzzy logic can be used to smooth the balancing between growing queues at ramps and traffic flow along the freeway.

In conclusion, beside the limited description of the gating techniques with pre-signals, no research paper was found dealing comprehensively with bimodal transportation management along congested signalised arterials.

4. THE PHILOSOPHICAL FRAMEWORK OF BIMODAL TRANSPORTATION MANAGEMENT

4.1 Introduction

This chapter discusses the existing philosophical framework that governs the development of urban transportation systems. It explores the ‘predict and provide’ approach and its limitations, the interdependences of long-term planning and short-term operations objectives, and the importance of the relative transit travel time improvements.

4.2 The Need for a New Philosophical Framework

The implementation of the bimodal transportation management concept is very difficult, if not impossible, within the traditional philosophical framework of transportation engineering and planning. To change current car oriented urban developments and the prevailing “cars-first” thinking and design practices, the traditional philosophical environment that has been the foundation of the transportation profession for many years needs to be revisited and modified.

4.2.1 Traffic Management vs. Traffic Accommodation

The required future roadway capacity is traditionally determined by predicting future use based on past trends and anticipated population and development growth. This approach, often called ‘predict and provide’, seems logical and it is well entrenched in the transportation profession.

The application of the 'predict and provide' in North America goes back to the 1926 Boston Transportation Study which used a simplified form of the gravity model to forecast traffic volumes (Weiner, 1992). During the 1940s, the relationship between land use and travel had been studied and sometimes used for travel forecasting (Weiner, 1992). However, until the 1950s traffic forecasting was done mostly by extending past traffic growth rates into the future (Weiner, 1992).

Modelling of future travel started in the mid 1950s with the Frater's technique (Schimdt et al. 1956) which was based on origin-destination data and growth factors, and the enhancements of the gravity model (Voorhees, 1956). The four-stage planning model (including trip generation, trip distribution, modal split, and traffic assignment) was formulated when the 1962 Federal-Aid Highway Act specified planning requirements for projects using federal funds (Weiner, 1992). This was the beginning of the "golden" era for transportation infrastructure in the US when the implementation of the 41,000 miles National Interstate Highway program was in full swing. Although similar massive roadway infrastructure building did not occur in Canada due to its sparse population, the four-stage planning model was adapted by the Canadian transportation profession as well.

In 1976, the Institute of Transportation Engineers (ITE) published the first edition of the Trip Generation report which included information on the number of vehicle trips generated by different land uses in various US (mostly suburban) areas. Since its first edition, the report has been updated six times (the 7th edition was published in 2003) and

it became the foundation of all transportation/traffic impact studies (TIS) on both sides of the border.

The wide acceptance of the four-stage planning model and of the various TIS procedures developed by individual municipalities and ITE guaranteed that the 'predict and provide' approach became the underlying principle at both the long-term regional planning level and the shorter-term development accommodation level. Although the 'predict and provide' approach seems logical and worked reasonably well for a number of years, it has limitations.

In the United Kingdom, due to the more dense urban structure, Buchanan recognized the limitations of building sufficient roadway capacity in urban areas much earlier than in North America. In his seminal book *Traffic in Towns* (Buchanan, 1963) predicted the need to curb the growth by writing: "Unless steps are taken, the motor vehicle will defeat its own utility and bring about a disastrous degradation of the surroundings for living... Either the utility of vehicles in town will decline rapidly, or the pleasantness and safety of surroundings will deteriorate catastrophically – in all probability both will happen."

Although Buchanan's prediction in terms of the negative impact of traffic on its surrounding environment has been evident, the increasing congestion and lack of alternatives forced the UK Government to continue with adding capacity to reduce congestions. One of the major initiatives was the Roads to Prosperity program in 1989 (Goodwin, 1997). However, in the view of Goodwin, the program failed very fast: "This

was the last time when any Government policy tried –even partially, and with caveats and exceptions – to devise a roads programme intended to ‘meet demand’... but even by the time of the launch the process which would lead to its abandonment was under way...The flaw was, the programme would not keep pace with traffic growth... even a fantasy road programme would not keep pace with traffic growth” (Goodwin, 1997).

In 1997, the UK Labour Government announced a moratorium on road building but five years later, due to the increasing congestion, it announced plans for new roadways and plans for roadway charging. One of the major initiatives has been the widening of the ring road around London, the M25 (BBC News, 2003). The construction of new capacity has been progressing very slowly and at a very high cost, with recent estimates of 21 million pounds per km (Jowit, 2007).

In the United States, as building the National Interstate Highway System drew to its completion in 1992 (Stufflebeam, 2004), it became obvious that the vehicle-miles travelled (i.e. the demand for road space) was growing much faster than the provision of roadways under the most ambitious public works project of highway building in US history.

The post-interstate era started with the 1991 Intermodal Surface Transportation Efficiency Act. The Act started to change the focus of roadway building only approach to the need for focus on other modes of transportation and on the better utilization of the infrastructure (Weiner, 1992).

In Canada, the importance of transit and transit oriented planning was always more pronounced than in the US. Today, all major cities place much higher priority in their Official Plans and/or Transportation Master Plans on alternative modes than on the car mode (Vancouver, 2007; Calgary, 2005; Toronto, 2007, Ottawa, 1997; Ottawa, 2003). The support for new roadway infrastructure has been gradually declining.

The overall decrease of interest to provide more roadway infrastructure at both sides of the Atlantic has at least three main reasons. First, it has been increasingly accepted, at least on the long-term planning level, that it is impossible to continue and/or accelerate road building programs that will keep pace with the growth in demand assuming the existing car-dependent transportation system. Second, public opposition makes it increasingly difficult and expensive to add significant roadway capacity in many urban areas. And third, environmental impacts and concerns, and the increasing acceptance of the need to reduce greenhouse gas emissions from transportation sources raise serious ethical and legal questions and concerns.

While building new roadways is decelerating, demand for travel continues to increase due to increasing population, car ownership, low density developments and the accompanying life styles. Measures to decrease demand for travel through Transportation Demand Management (TDM), pricing, or to accommodate travel demand by more efficient modes such as transit have not yet been sufficiently successful. The 'predict and provide' approach became a 'predict and under-provide' approach (Goodwin, 1997).

It is obvious that the discrepancy in the demand and supply of roadway space creates congestion. The awareness of the unsustainability of the 'predict and provide' approach and the society's inability and/or unwillingness to implement effective changes creates an uncomfortable equilibrium. Owens (1995) considers this uncomfortable equilibrium to be a paradox. Congestion has a price, but this price in most urban areas is not yet high enough to trigger fundamental changes in urban design, the provision of transportation facilities and modes, and in lifestyles. Therefore, it is likely that most major urban areas will continue to face congestion for a significant number of years.

This is the traffic and political environment in which bimodal traffic management can still be applied relatively easily from the technical perspective. This is the environment in which the bimodal thinking should be provided because it improves the overall efficiency of the transportation system and triggers the shift from car travel to transit travel.

But the technical solutions cannot happen in isolation from the philosophical/planning background. If the underlying assumption remains to be the 'predict and provide' approach irrespective of its limitations, and the assumption of limitless capacity is maintained, bimodal traffic management will remain on the fringes.

4.2.1.1 The Limitations of Roadway Capacity Increase

Planning models and approaches generally assume no vehicular capacity limitations. For example the two most frequently used planning models, the four-stage planning model

and the transportation impact assessment model assume by design that the potential for vehicular capacity increase is practically limitless.

The four-stage planning model is regularly used in long-range planning to forecast the required roadway and transit infrastructure capacity. Transportation Impact Assessments are required by municipal by-laws and are used to support development applications by assessing the shorter-term roadway infrastructure needs. Transportation Impact Assessment guidelines are normally developed by municipalities (Ottawa, 2006; Toronto, 2003; Peel, 2005; Alberta, 2005, Florida DOT, 1997) and although they differ from one municipality to another, they generally follow the ITE Transportation Impact Assessment Guidelines (ITE, 2006).

The common element of both planning procedures is that they forecast the requirement for vehicular capacity, and the outcome of the analysis process is the number of required additional vehicular lanes, which is the 'predict and provide' approach.

If there are capacity limitations and the required additional lanes cannot be all provided, the situation is normally treated as a temporary anomaly, and additional roadway capacity is provided to the maximum extent. This approach is somewhat similar to filling up the fuel tanks of an aeroplane only 50% because of the shortage of fuel, even though the flight requires full tanks.

A new philosophical framework for transportation planning models is proposed (Figure 4.1). In circumstances when vehicular capacity restrictions exist, the capacity restriction

becomes, by design, an input into the planning model. The outcome of the model is the required transit modal share (and/or pedestrian/cycling modal share). This approach would formalize, support, and trigger the development of physical and policy measures that have to be implemented to achieve the required modal shares. Instead of resisting the shift to more sustainable modes, the new philosophy would promote it.

In terms of the bimodal transportation management concept, the proposed framework would provide the required support for using existing or the limited additional roadway capacity for improving transit service or other non-car modes, instead of using it up for marginal vehicular capacity benefits.

The idea of the proposed shift in the philosophical framework is not entirely new. Owens (1995) writes about the possibility of changing the 'predict and provide' approach to a 'predict and prevent' approach. A similar approach, the concept of 'environmental capacity' was originally developed for preventing marine pollution (GESAMP, 1986) and it is also being explored in land use planning (Rydin, 1998) and in air transportation (Upham, 2001). The basic principle is to determine a sustainable capacity of the natural or man made systems, and to adapt to it.

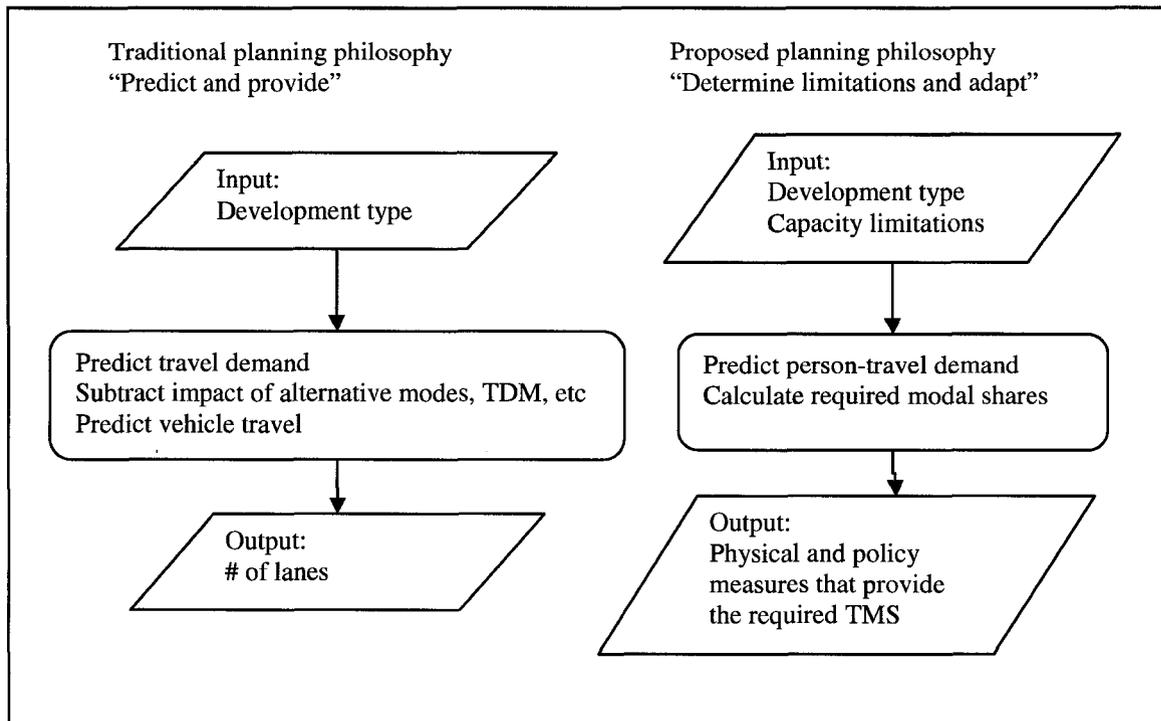


Figure 4.1: The Comparison of the Traditional and the Proposed Planning Philosophy

One of the main issues of the 'environmental capacity' approach is how to determine the capacity limitations. In case of natural systems, the complexity stems from the difficulties of measurement, the lack of knowledge of potential consequences, and the subjective perception of consequences. In man made systems, such as the transportation system, determining the 'environmental capacity' has its own challenges which include the possibilities of new technologies and the subjective nature of assessing some of the benefits/disbenefits. Nevertheless, it appears that determining the capacity limitations of transportation systems, is still much simpler than is the case for natural systems.

4.2.2 The Connection between Long-Term Transportation Planning and Short-Term Operations Objectives

In the traditional philosophical environment, the transportation operations function is rather isolated from the transportation planning function. For example, in the transportation planning domain, the unsustainability of car-oriented low density developments is recognized. Transportation planning and policy directions, through long term Transportation Master Plans (Ottawa, 2003), are promoted which seek to discourage this type of development. However, the same documents are also promoting the provision of sufficient roadway infrastructure to support operational objectives in the form of “adequate” roadway capacity. The problem is not with the provision of “adequate roadway capacity”, it is rather the absence of policy approaches/directions when both objectives cannot be achieved due to resource (i.e. right-of-way, financial) limitations. In these circumstances, the unavoidable trade-offs are made somewhat unpredictably at the political level by leaving the transportation profession on the sideline.

To increase the input of the transportation profession in the decision making process, the Transportation Planning and the Transportation Operations functions must be coordinated and mutually supportive. It is expected that this coordination can be achieved by developing MoP that are used by both planning and operations. Further discussion on these MoP is provided in Chapter 5.

4.2.3 Improvements to All vs. Relative Transit Improvements

The problem of bimodal urban transportation management is the Prisoners' Dilemma (Stanford, 2008) in disguise. The Prisoner's Dilemma translated to commuters is the following:

If all commuters travelled by bus ('prisoners cooperate'), they would all spend time walking to/from the stop and waiting (which represents the out of vehicle travel time - OVTT). However, they would not be delayed by congestion (Scenario A, Table 4.1). The other extreme is when all commuters travel by car (Scenario C). They would not experience any OVTT, but they would suffer from congestion. Depending on congestion delay and the length of OVTT, it is quite possible that the mean travel time is shorter if all commuters travelled by bus. Smeed et al. (1964) explored this possibility and based on data analysis for central London concluded that if all commuters travelled by bus, the mean travel time from door to door for all passengers would slightly decrease.

Table 4.1: The "Commuter's Dilemma"

Scenario	Action of commuter		Travel Time of commuter		Mean travel time for all commuters
	A	B	A	B	
A	Bus	Bus	OVTT no congestion delay	OVTT no congestion delay	OVTT but no congestion delay
B	Bus	Car	OVTT and congestion delay	a bit shorter congestion delay	depends on the % of bus riders
C	Car	Car	highest congestion delay	highest congestion delay	highest congestion delay but no OVTT

However, as Goodwin (1997) pointed out, as long as buses travel in the same traffic stream with cars, there is no incentive for car drivers to switch to transit. The individual's trip that continues to travel by car is always shorter than the bus riders trip (Scenario B). So the car driver faces a dilemma. If he switches to transit and the other car drivers don't he is worse off. If the other car drivers switch as well, they are all better off.

In real-life, the commuters' dilemma is relatively simple. The travel time benefits of driving compared to travelling by bus are much more certain than the remote possibility of many other motorists to make the switch. Therefore, expecting, that motorists will change modes because that would be better for everyone on average, is probably too naïve. To move towards the system optimum (Scenario C), where the modal share of transit increases, the traffic control and the transportation infrastructure must allow transit vehicles to travel without congestion delay. Therefore, the provision of a relative advantage to transit compared to car travel times is required to achieve optimum system operation.

4.3 The Foundations: Why does Bimodal Queue Management Work?

To facilitate uncongested transit vehicle travel through a congested network (without grade separating the two) the control system has to provide two functions, the queue management function and the queue jump function for transit vehicles.

The queue management function developed in this thesis has the following two general objectives:

- to move the queue formation to a location where the queue causes the least “damage”;
- to have no or minimal impact of vehicle capacity and total vehicle travel time.

The first objective is always location specific. It requires the definition of “damage”. For the purpose of this thesis, the definition of damage is limited to the negative impacts on transit vehicle movement.

The feasibility of the second objective is discussed through the following example:

System delay on a transportation network comprise two major components; delay to enter the network and delay to traverse the network (Figure 4.2). Delay to entry is the time spent before the user can enter the network (e.g. delay at the ramp meter signal). Delay in the network is the time spent traversing the network.

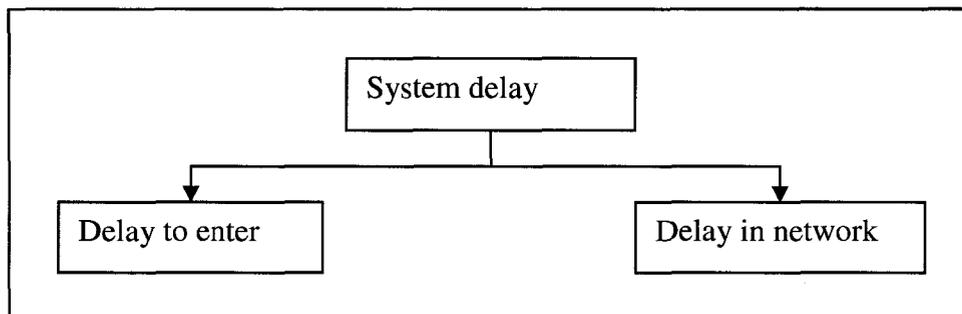


Figure 4.2: The Components of System Delay

The traditional traffic engineering practice is to minimize both the *delay to enter* and the *delay in the network* independently. This approach is often sensible and will always provide minimum system delay (total time in the system) in unsaturated conditions.

In congested conditions, however, a different strategy may provide the lowest system delay. When an arterial or network becomes oversaturated, its vehicle capacity may get reduced rapidly due to the queue blockages described in Section 2.2.1.1. In those circumstances lower total system delay may be achieved by metering the volume at the entry point(s). By controlling the number of vehicles in the system, queue blockages and the associated capacity losses and rapid travel time increases could be minimized or eliminated. Roess et al. (1998) describes three metering strategies:

- internal metering;
- external metering; and
- release metering.

Internal metering breaks up the queue on the approach leading to the bottleneck intersection. This is achieved by limiting green time at upstream intersections to a degree which prevents long queue formation at the critical intersection approach (Roess 1998).

External metering is done at the peripheries of an area by limiting the number of vehicles entering at strategically selected control points (Roess et al., 1998). Release metering reduces the peaking characteristics of discharge flows from large traffic generators.

Although volume metering has a strong theoretical appeal, its implementation in traffic control (except ramp metering) has been rather limited. In other areas, both in the technical field and in human activities, the metering concept is utilized extensively. For example in the distribution of electricity, if demand exceeds capacity, some consumers are denied service instead of allowing system break down; party goers are held back at

the door until someone leaves when the dancing floor has reached its “capacity”; the washing machine is not filled up 100% of its capacity, the extra laundry is left to wait for a second load, otherwise the laundry will remain dirty (Joos, Undated).

There are a number of reasons why the metering concept has not yet been developed for arterial/network traffic management:

- The Traffic Engineering profession has been focusing on resolving the congestion by increasing capacity. This is despite the long-term planning opinion that the increase of the roadway supply cannot match the increase in vehicle demand with traditional car-oriented land use development.
- Effective metering must be based on dynamic control, which is significantly more complex than most signal control systems. Beside the technical complexities, public perception, potential diversions, and equity issues may also arise.
- Reducing total vehicle delay by metering may be difficult to achieve or/and prove unless the metering scheme impacts a small and well defined section of the arterial/network. Field measurements may be too complex to be practical and reliable. Microscopic simulation is a potentially good tool, however, it has also number of challenges.

All the above three reasons are addressed in this thesis. The connection between Planning and Operational objectives and the required performance measures are discussed in Chapter 5. Dynamic control techniques, are explored in Chapter 7, Chapter 8, and

Chapter 9. By combining queue management with transit queue jumps, the benefit of queue management becomes much more obvious.

4.3.1.1 The Transit Queue Jump Function

The objective of the queue jump function is to change the sequence of vehicles in the queue by moving high capacity vehicles (transit vehicles) to the beginning of the queue. In queuing theory, this is described as the transformation of a FIFO (first-in-first-out) system into a LIFO (last in first out) system.

By changing the sequence of vehicles from FIFO to LIFO, total vehicle delay does not change, as long as the service time of each vehicle is the same. However, if a vehicle, with high occupancy goes from the end of the queue to the beginning of the queue, total person travel time (or delay) decreases. This is explained on an elementary example.

Figure 4.2 shows two alternative vehicle sequences at an intersection approach. Assume that vehicle B (the high occupancy bus) in both alternatives arrived last, but in Alternative B, through special geometric set-up (not shown here) gets to the beginning of the queue.

At this stage, only the start up delays are considered in the evaluation. The wait time for the first green phase is the same for both alternatives, thus it is ignored. Assume that the service time of cars (gap between cars entering the intersection) is 2 seconds (saturation flow rate=1800 veh/hr) and the service time of a bus is 4 seconds due to its size. To

service all 10 vehicles and the bus, both alternatives need the same amount of green time, 26 seconds (Table 4.2).

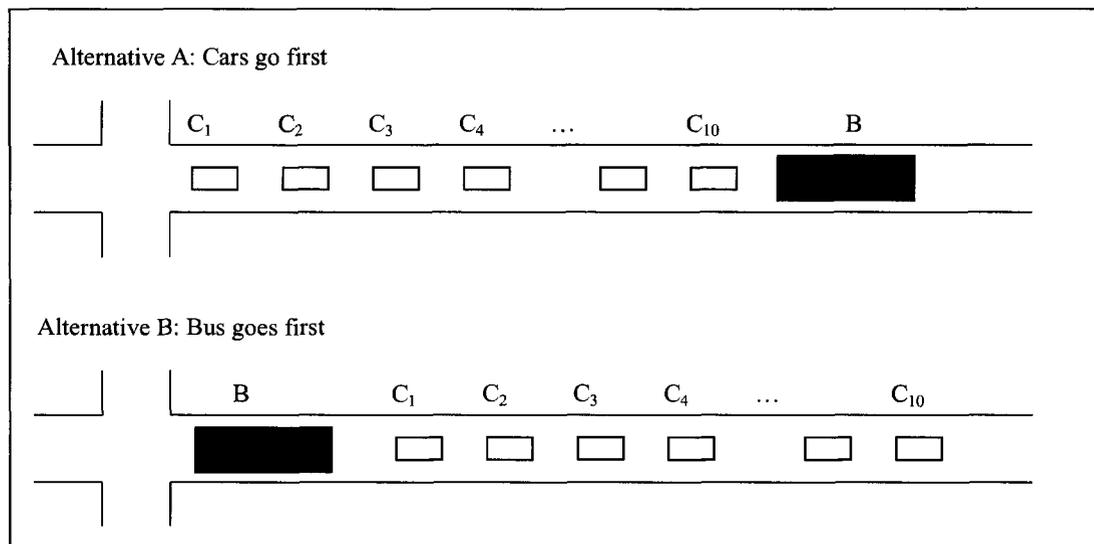


Figure 4.3: Alternative Priorities at an Intersection Approach

In terms of total vehicle delay, Alternative A is slightly better. Alternative A generates 132 seconds total vehicle delay while Alternative B generates 152 seconds. Alternative B has higher delay, because the transit vehicle is serviced first and due to its size it requires more time to enter the intersection (longer service time), and this additional time is added to all cars that follow the bus.

From the person delay perspective, Alternative B is significantly better. When the bus goes first, person delay drops to almost half compared to when the bus goes last. (Bus occupancy is 10 person/veh and car occupancy is 1 person/veh.)

The evaluation generated two contradicting results. If total-vehicle delay (or mean vehicle delay) is selected as the observed MoP, the alternative without a queue jump is better. This implies that the approach will have the least total vehicle delay if vehicles that require the longest service time are serviced the last.

Table 4.2: Conceptual Analysis of MoP

Alternative A: Without Transit Queue Jump			Alternative B: With Transit Queue Jump		
Vehicle	Time of departure (sec)	Person delay (sec)	Vehicle	Time of departure (sec)	Person delay (sec)
C1	2	2	B	2	20
C2	4	4	C1	6	6
C3	6	6	C2	8	8
C4	8	8	C3	10	10
C5	10	10	C4	12	12
C6	12	12	C5	14	14
C7	14	14	C6	16	16
C8	16	16	C7	18	18
C9	18	18	C8	20	20
C10	20	20	C9	22	22
B	22	220	C10	24	24
Total green time	22+4=26		Total green time	24+2=26	
Total delay	132 veh-sec	330 pers-sec	Total	152 veh-sec	170 pers-sec

If total person delay is selected as the critical MoP, the alternative with the high occupancy vehicle (bus) is being served first is significantly better.

Traditionally, vehicle delay has been one of the most frequently used MoP for design, control, and evaluation purposes. However, through the previous examples it has been shown that if the design and operation of the transportation system has to achieve its

planning objectives, a philosophical paradigm shift in traffic engineering must occur as well. A MoP that reflects the person focus must be selected as a fundamental MoP.

Further discussion on the MoP is provided in Chapter 5.

4.3.2 Discussion and Conclusions

The concept of queue management is, to a certain degree, in contradiction with the inherited human (especially) engineering approach to problems: we do not like limitations, and we do not like trade-offs. However, the concept of queue management requires the recognition of limitations to our technical/societal ability to provide sufficient vehicular capacity, and it also requires to deal with the unavoidable trade-offs.

The implementation of the bimodal transportation management concept is hampered by the existing philosophical framework of transportation engineering and planning. This framework has to be modified in order to promote change instead of resisting it. Building of roadway infrastructure has been traditionally based on the 'predict and provide' approach. However, in congested urban environment, the approach became 'predict and provide as-much-as-possible'. This approach however, is not too beneficial. The development of a 'determine capacity limitations and adapt' approach is proposed.

A probably surprising aspect of some bimodal queue management techniques is that dealing with trade-offs may not be too difficult after all. It is possible to demonstrate that the person travel time to all users at an intersection approach can be reduced by changing

the sequence of vehicles in the traffic stream without incurring a significant increase in mean vehicle travel times.

5. BIMODAL PERFORMANCE ANALYSIS FRAMEWORK

5.1 Introduction

This section describes the development of MoP proposed for the evaluation of bimodal transportation management strategies in congested conditions. The section starts with establishing the connection between long-term transportation policy objectives and MoP. This is followed by a description of the most widely used traditional MoP and a discussion of their limitations in bimodal operation in congestion. Finally the proposed MoP and their limitations are presented.

5.2 The Connection between Objectives and MoP

Public investments in transportation facilities, innovations in traffic control and management are normally attempts to achieve transportation policy/planning objectives and/or vision. This vision has a long-term focus and its primary role is to encourage and support the development and implementation of a series of short-term actions that will eventually achieve the long-term policy/planning objectives (Figure 5.1). This vision is at the first (highest) level of planning and is described in linguistic terms such as “enhance our quality of life” (Ottawa, 2003), or create a “modern transportation system” (Brampton, 2003), and it is spelled out in relatively vague terms. To achieve long-term policy/planning objectives, design objectives are formulated at the second level to support transportation system planning and design (Figure 5.1). System design objectives

are often quantified by MoP. At the third level, operations/control objectives are formulated and quantified through MoP.

In an ideal world, the objectives at all three levels (vision, design, and operations) are mutually supportive. Furthermore, the MoP at different levels are either identical or the relationship between them is transparent and well understood. The single-mode unsaturated conditions (SM-US) scenario is very close to the ideal situation.

For example, at the vision level, the objective can be (and has been) defined as improve (vehicle) mobility, support (car based) development, enhance (vehicle) connectivity and so on. At the design level, the objectives are pretty straightforward: increase vehicular capacity. The MoP describe the adequacy of the new capacity (V/C) and/or the design performance expressed in speed, travel time or delay. At the operations level, the objectives are still obvious, the control system has to provide the least possible delay, and the same MoP can be used as at the design level (V/C and delay).

The mutually supportive connection between various levels of objectives and the similarity in the MoP makes the implementation of SM-US systems very attractive.

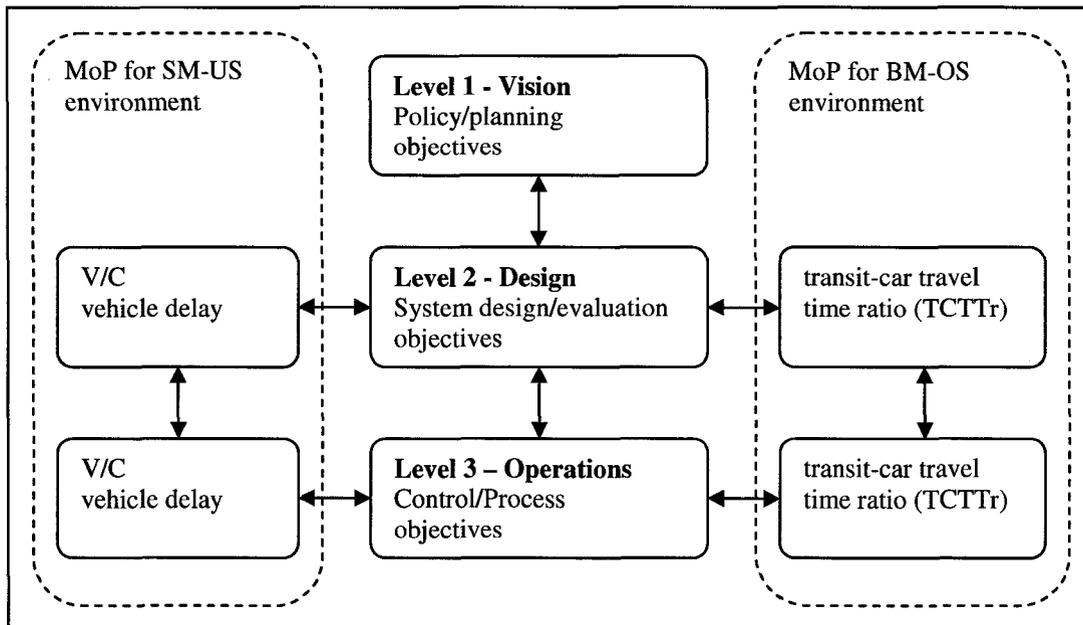


Figure 5.1: The Relationship between Objectives at Various Levels and MoP in Single-Mode Unsaturated and Bimodal Oversaturated Conditions

The development of a transportation system for oversaturated (BM-OS) conditions is much more difficult for the following reasons:

- potentially conflicting policy/planning objectives at the Vision level; and
- objectives at the Vision, Design, and Operations levels may not be mutually supportive, and the lack of a comprehensive traditional MoP.

Potentially conflicting policy/planning objectives

Visions usually reflect aspirations and not the implementation difficulties. Therefore, the long-term policy objectives in a MM-OS environment are multi dimensional with potentially contradicting objectives. In a constrained environment, policy objectives such as reducing congestion, increasing transit modal share, and providing for growth are likely contradicting objectives.

For example, there is a potential trade-off between decreasing congestion by increasing vehicular capacity and increasing transit modal share. Expanding the roadway may help reduce congestion and provide for growth, but it also contributes to a more car oriented land use which does not support the increase in transit ridership. On the other hand, improving transit may help to provide for growth but may do little to reduce existing vehicular congestion. The potential conflicts must be resolved at the Vision (policy/planning level) to facilitate the trade-offs at the Design level.

Objectives at the Vision, Design, and Operations levels may not be mutually supportive

If the potential trade-offs at the Vision level are not resolved, the definition of Design and Operations level objectives are problematic. For example, the policy/planning objective to reduce vehicular congestion can be interpreted at the Design level as either to increase vehicular capacity or to improve transit travel times. At the Operations level, the Vision can be interpreted as a direction to improve vehicle travel times by reducing intersection delay or to improve transit travel times relative to car travel times.

For example, the Vision level objective is to increase transit ridership. At the Design level, the traffic control is enhanced to provide selective priority to transit vehicles at intersections by keeping the green signal longer when transit vehicles approach the intersection. However, the maximum duration of green extension i.e. level of priority is normally an Operations level decision. If the Operations level objective is to minimize

vehicle delay (vs. improve relative transit travel times), the objectives at the three levels are mutually not supportive.

Transportation related decisions are significantly influenced by short-term public opinion and political opportunities. Therefore, it should not be surprising when the long term planning objectives contain contradicting elements or when the various levels of decisions are not mutually supportive. This climate of influence is not expected to change.

However, there are possibilities to minimize the impacts of these conflicts by clearly identifying them. It is proposed, that the selection of the MoP could be an effective way to “enforce” the coordination of objectives at the Vision, Design, and Operations levels. One way to achieve this is by developing/selecting MoP than can be applied at all three levels.

For example, it is known that transit modal share is significantly influenced by the *relative* advantage of transit travel times compared to car travel times. Therefore, a Vision level objective that has a transit focus (which is the case in most Canadian urban environments) must spell out the need for *relative* improvements in transit travel times. At the Design level, the *relative* improvement objective can be transformed into infrastructure design elements, and at the Operations level the control objectives would be clarified.

There have been attempts to steer the development of the transportation system by specifying target MoP. For example, the City of Ottawa has set a 30 % Transit Modal Share (TMS) target in its Official Plan (Ottawa, 2003). However, the TMS target does not represent a target at the Design or Operations level decisions. Although the TMS target remains an aspiration goal for high level decision making and a benchmark for comparison purposes, Design level decisions remain to be influenced by traditional MoP which have a vehicle focus.

In Chapter 5.3 the traditional MoP will be further examined and in Chapter 5.4 recommendations are made regarding the most appropriate multi-modal MoP. Before that, however, a paradox related to the selection of MoP is presented.

5.2.1 “Gaining Efficiency by Abusing Efficiency Paradox” and the “Propagation of Inefficiencies Trap”

The Paradox and the Trap are two sides of the same coin. They both describe the same situation but from somewhat different perspectives.

Suppose the objective is to optimize system operation by criteria A. If a specific element of the system is efficient from the criteria A perspective, improvements in system performance may be achieved by abusing the efficient elements and/or compensating the inefficient elements. In other words, because inefficient elements use most of the resources, they should be prioritised in the allocation of resources in order to minimize resource use. This is the “gaining efficiency by abusing efficiency paradox”. Since the

efficient element of the system is “abused” and the inefficient element receives priority, the negative impact of inefficient elements will eventually increase. This is the “propagation of inefficiencies trap”.

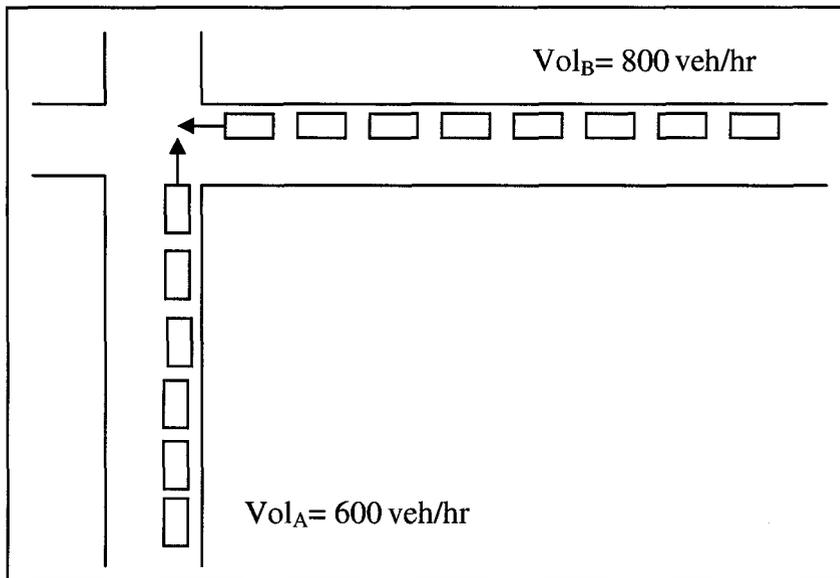


Figure 5.2: Example 1-Single-Mode Operation at Isolated Intersection

In congested conditions (Figure 5.2), a common principle to allocate green time between conflicting movements (split calculation) is to equalize the volume over capacity (V/C) ratios as described by Equation 5.1. The volume represents the number of vehicles approaching the intersection from one direction within a time interval, usually one hour, and capacity is the maximum number of vehicles that can enter the intersection. The capacity of an approach can be estimated by using Equations 5.2 and 5.3 for approaches A and B respectively.

$$\frac{V_A}{C_A} = \frac{V_B}{C_B} \quad 5.1$$

V_A, V_B – volume on approach A and B

C_A, C_B – capacity of approach A and B

$$C_A = n_A \times \frac{g_A}{cyc} \times s_A \quad 5.2$$

$$C_B = n_B \times \frac{g_B}{cyc} \times s_B \quad 5.3$$

n_A, n_B – number of lanes on approaches A and B

g_A, g_B – green time at approaches A and B

cyc – cycle length (sec)

s_A, s_B – saturation flow rate at approaches A and B (veh/hr)

If we assume that there is only one lane at each approach (Equation 5.4), the ratio of allocated green times for approaches A and B can be expressed by Equations 5.5.

$$n_A = n_B = 1 \quad 5.4$$

$$\frac{g_A}{g_B} = \frac{V_A}{V_B} \times \frac{s_B}{s_A} \quad 5.5$$

Equation 5.5 reveals that the allocated green time to approach A is proportional to the volume on approach A and inversely proportional to saturation flow at the same approach. A similar conclusion can be made for the other approach as well. The relationship ensures that the approach that is operating closer to its capacity, the more ‘problematic’ approach, gets more green time, i.e. priority. This logic, in congested conditions, ensures that the queues at conflicting approaches dissipate approximately at the same time.

The dissipation of queues approximately at the same time has at least two major benefits. First, it provides the perception of an equitable treatment of different intersection approaches. Second, it provides optimum system operation defined by the minimum total vehicle delay (Gazis et al., 1965). However, it is important to point out that this operation does not generate equal vehicle delay at the different approaches.

Vehicle delay can be calculated by using the Canadian Capacity Guide (CCG) (CITE, 2008), the Highway Capacity Manual (HCM, 2000), or other traditional delay formulas. For the purpose of this paper, the CCG method was used but similar results are expected by using any other delay formulas. (Delay formulas are further discussed in Section 5.3.1)

The assumed input values and the corresponding delay are summarized in Table 5.1. A cycle length of 120 sec, clearance intervals of 8 sec, same number of lanes and saturation flow rates, and two phase operation were assumed. Green times were calculated by using Equations 5.5. Delays were calculated by using the CCG method formulated by Equations 5.6, 5.7, and 5.8.

$$d = k_f \times d_1 + d_2 \quad 5.6$$

$$d_1 = \frac{cyc \times \left[1 - \frac{g}{cyc} \right]^2}{2 \times \left[1 - x_1 \times \frac{g}{cyc} \right]} \quad 5.7$$

$$d_2 = 15 \times t_e \times \left[(X - 1) + \sqrt{(X - 1)^2 + \frac{240 \times X}{C \times t_e}} \right] \quad 5.8$$

k_f – adjustment factor for progression (=1 for isolated intersection)

x_1 – minimum of (1.0, V/C)

Table 5.1: Green Time and Delay Calculation for Example 1

	Volume (veh/hr)	Saturation flow (veh/hr-green)	Green (sec/cyc)	Delay (sec/veh)
Approach A	600	1800	48	43
Approach B	800	1800	64	32
cyc=120 sec, clearance=8 sec, $t_e=15$ min, $k_f=1$				

The calculation results reveal that vehicles on the higher volume approach get longer green time, and the associated mean vehicle delay is less compared to the approach with less volume. This result is not surprising. The inequality of delay between the high and low volume approaches (low volume is delayed more) can be justified relatively easily by the achievement of system optimum in terms of total vehicle delay.

Figure 5.3 shows the impact of the difference in volumes of A and B on approach delays.

The bottom graph shows input volumes. The volume on approach A has been kept constant at 600 veh/hr, and volume on approach B has been increased from 200 veh/hr to 1000 veh/hour. The top graph depicts the corresponding mean delays for the two approaches. The x axes for both graphs are the same and indicate the volume over capacity ratio which is identical for the two approaches. At the left end of the graphs, Volume B is lower (bottom graph), thus delay on approach B is higher (top graph). As we move to the right, the difference in volume and delay on approaches A and B decreases.

At equal volumes on A and B, delay is equal. At the right end of the graph volume B is higher (bottom graph), and delay on approach B is lower (top graph).

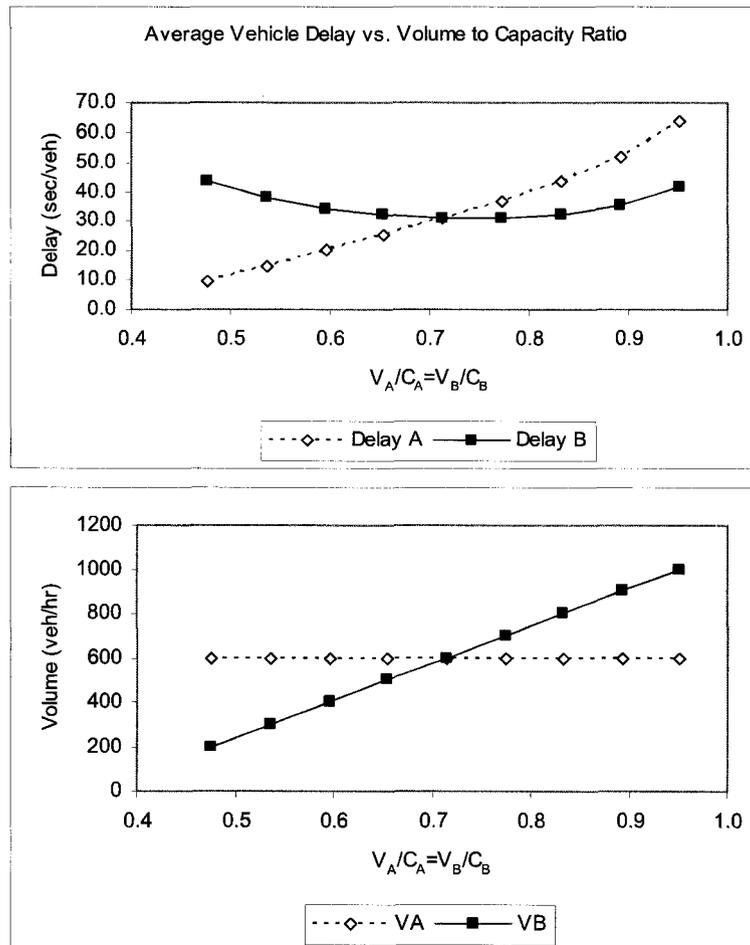


Figure 5.3: Sensitivity Analysis

The important thing that is revealed by Figure 5.3 is that the difference between the mean delay on approaches A and B increases with the difference in volumes on approaches A and B. As discussed earlier, this is justified from the system efficiency perspective

(minimum total vehicle delay is achieved). However, this relationship leads to a paradox in bimodal traffic control.

Let's consider the following bimodal example. The vehicular volume on approach B is identical to Example 1 but approach A carries 20 buses per hour (instead of 600 veh/hr). Assume a bus equivalency factor of 2, thus 20 bus/hr are equal to 40 passenger car units (pcu).

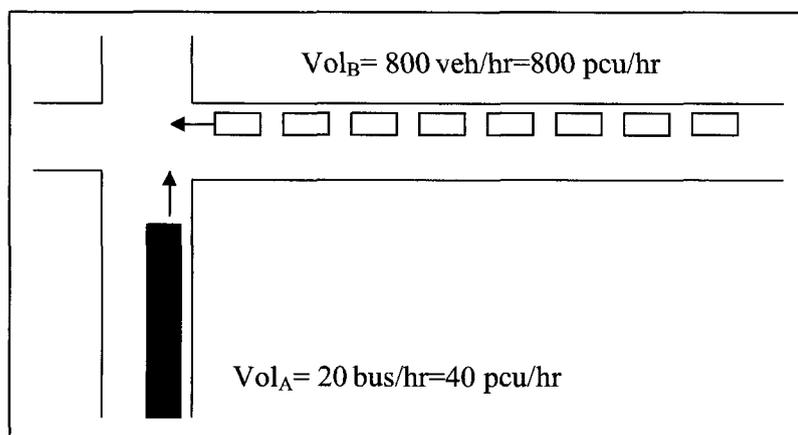


Figure 5.4: Example 2 – Bimodal Operation at Isolated Intersection

The green splits and the corresponding approach delays are calculated similarly to Example 1. The results are in line with previous observations. The higher volume approach gets more green time and the resulting mean delay is smaller than at the lower volume approach. The big difference between approach A and B volumes results in a big difference in mean vehicle delay between the two approaches. Approach A with 20 buses has an mean delay of 77 sec while the delay on approach B is only 3 sec (Table 5.2).

Table 5.2: Green Time and Delay Calculation for Example 2

	Volume (pcu/hr)	Saturation flow (veh/hr-green)	Green (sec/cyc)	Delay (sec/veh)
Approach A	40	1800	5	77
Approach B	800	1800	107	3
cyc=120 sec, clearance=8 sec, te=15 min, kf=1				

While in Example 1 the justification of unequal delays for the two approaches seemed to be reasonable, the same rationale for Example 2 needs further scrutiny.

The reason why the system optimum requires only 5 sec green time for approach A is because the approach operates much more efficiently than approach B. If one bus carries 30 people, the 20 buses on approach A carry the same number of travelers as in Example 1. But now (Example 2), approach B needs only 5 sec green time compared to 48 sec (Example 1). Since all the extra green time is allocated to approach B, the mean delay for the 800 vehicles on approach B goes down to only 3 sec.

The equalization of the V/C achieved system optimum for both Example 1 and 2.

However, system efficiency has been achieved by reducing the delay for the inefficient element and by increasing delay to the efficient element. This represents the “gaining efficiency by abusing efficiency paradox”.

The traditional traffic control/management logic in congested conditions leads to the same paradox in many situations. Let's consider an arterial operation depicted in Figure 5.5.

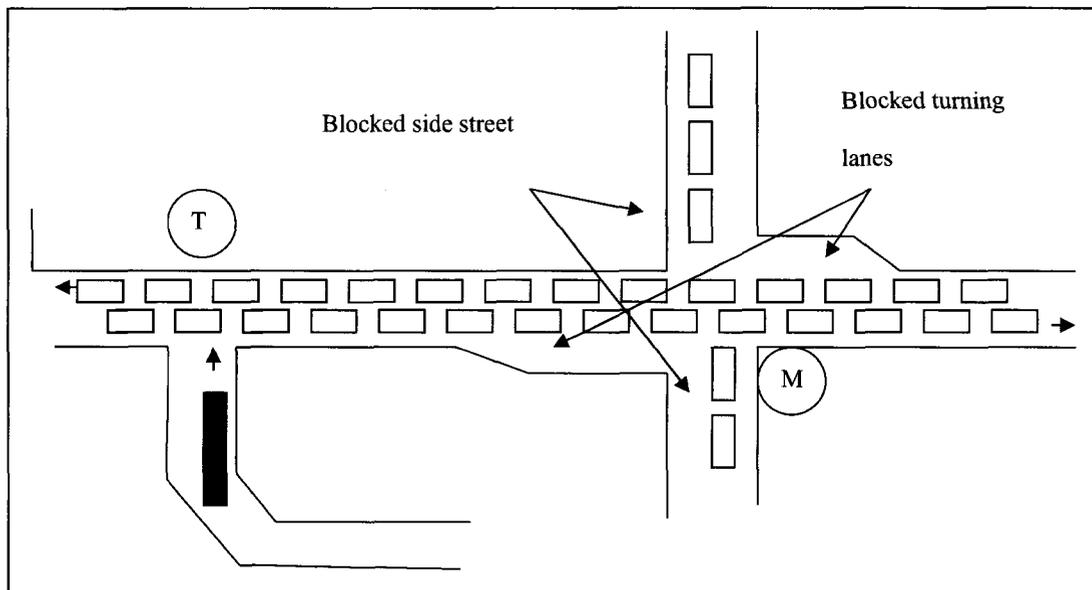


Figure 5.5: Example 3 – Arterial Operation

There is a long queue in the East-West direction which from time to time blocks side street traffic and impedes access to turning lanes at intersection 'M'. There is also a bus terminal connection 'T' to the arterial. The objective of traditional control is to minimize total vehicle delay which can be achieved by equalizing the V/C ratio. The outcome of the optimization is a significantly longer green time for the East-West movement at intersection 'T' than to the bus approach.

The importance of providing as much as possible green time to the East-West approach is elevated by the lane blockages at intersection M. Lane blockages at 'M' reduce the North-South discharge flow rate, delays right (and left) turning vehicles on the East-West approach. The delay created by these blockages is the secondary impact of congestion and it is in addition to the delay described earlier by the delay formulas.

The perception of the problem, the engineering judgement, and the mathematical optimization model all support the V/C equalization approach. The duration of congestion (time required to discharge the East-West queue) will be the shortest if the East-West flow is prioritized or interrupted as little as possible. From the congestion management perspective it is totally irrelevant how many people travel on buses. Travellers on the bus are 'stored' very efficiently and to reduce queue lengths and to minimize the duration of congestion, car travel must be prioritized. The "gaining efficiency by abusing efficiency paradox" applies.

One may argue that although a paradox exists, there is not much the transportation profession can or should do about it. So let's examine another somewhat different but conceptually identical example.

The objective of intersection control is to minimize air pollution (Figure 5.6). Let's suppose that vehicle A is efficient from the air pollution perspective – it emits low level of pollutants, while vehicle B is inefficient because it emits much higher levels of pollutants. To optimize system operation e.g. minimize total pollution emission, the inefficient element, vehicle B, should be "compensated" e.g. serviced first so that it spends the least time in the system. In other words, the efficient element, vehicle A, should be "abused" (receive lower priority) since its contribution to the system emissions is much less.

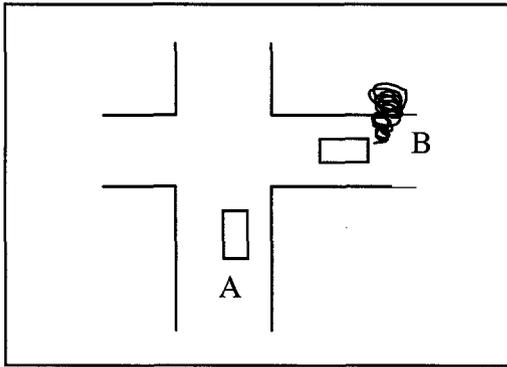


Figure 5.6: Example 4 – Minimization of Pollution

This example somehow suggests, more than the previous examples, that the logic of achieving system optimum by providing priority to the polluting (less efficient) vehicle has a flaw: The lowest total emissions are achieved by ‘abusing’ low polluting vehicles and the long term impact of such optimization logic is likely to increase the number of polluting vehicles. This is the “propagation of inefficiencies trap”.

Interestingly, this trap is usually much less obvious in the first three examples. The connection between the “gaining efficiency by abusing efficiency paradox” and the “propagation of inefficiencies trap” is not recognized at all in traffic control and management. The provision of more priority to the less efficient mode (in terms of road space and green time utilization) is the norm. Currently there is no traffic control optimization model that incorporates the long-term impact of providing higher priority to lower efficiency vehicles (cars).

The “gaining efficiency by abusing efficiency paradox” is an unavoidable challenge in bimodal transportation system optimizations. Optimum operation (in terms of vehicle

delay) is achieved by prioritizing vehicles that contribute the most to queue formations (and congestion). However, specific steps can be taken to avoid the trap.

The first step in avoiding the trap is to recognize it. The second step is to select MoP that capture all relevant characteristics of the system. The third step is to acknowledge the existence of relevant characteristics/relationships even if they cannot be measures and express them qualitatively.

5.3 The Applicability of Traditional MoP in Bimodal Congested Conditions

The Traffic Engineering practice utilizes a number of MoP in system design, evaluation and control. One of the most thoroughly researched and most widely applied MoP is the mean vehicle delay. It has been already suggested that vehicle delay is a poor performance measure of multi-modal systems. However, vehicle delay has a prominent place in most Traffic Engineering theories and practices thus before one could suggest that this MoP is not a good measure for bimodal assessment in congested conditions one must explore and expose its relevant weaknesses.

5.3.1 Mean Vehicle Delay

Vehicle delay models were developed as a MoP by a number of researchers. As described in Figure 5.1 vehicle delay is used as a performance measure at both the design/evaluation level and at the control/process level. At the design/evaluation level, vehicle delay is used to design intersection geometry and signal timing to achieve a

specific performance target. At the control/process level, vehicle delay is used in real-time signal operation (e.g. actuated operation, real-time system such as SCOOT).

5.3.1.1 The Calculation of the Mean Vehicle Delay

There are number of delay models used for signalized intersections including the Australian –Akcelik model (Akcelik, 1993), the Canadian Capacity Guide model (CCG) (CITE, 2008), and the Highway Capacity Manual (HCM) model (TRB, 2000). The three models are very similar and they are all based on Webster’s delay formula with enhancements to cover saturated and oversaturated conditions.

Webster’s delay formula published sixty years ago (Webster, 1958) has three components (5.9). The first component d_1 (5.10) describes uniform delay assuming a constant arrival rate and is represented by the area between the cumulative arrival and departure curves in Figure 5.7.

$$d = d_1 + d_2 - d_3 \quad 5.9$$

$$d_1 = \frac{cyc \times \left[1 - \frac{g}{cyc} \right]^2}{2 \times \left[1 - \frac{g}{cyc} \times X \right]} \quad 5.10$$

$$d_2 = \frac{X^2}{2 \times q \times (1 - X)} \quad 5.11$$

$$d_3 = 0.65 \times \left(\frac{cyc}{q^2} \right)^{\frac{1}{3}} \times X^{2 + 5 \times \frac{g}{cyc}} \quad 5.12$$

d – mean delay (sec/veh)

d_1 – mean uniform delay (sec/veh)

d_2 – mean random delay (sec/veh)

d_3 – correction (sec/veh)

cyc – cycle length (sec)

g - green time per cycle

q – arrival flow rate (veh/sec)

C - capacity (veh/hr); $C=(g/cyc)*s$

s - saturation flow rate (veh/hr)

V – volume (veh/hr); $V=q*3600$

X – degree of saturation; $X=V/C$

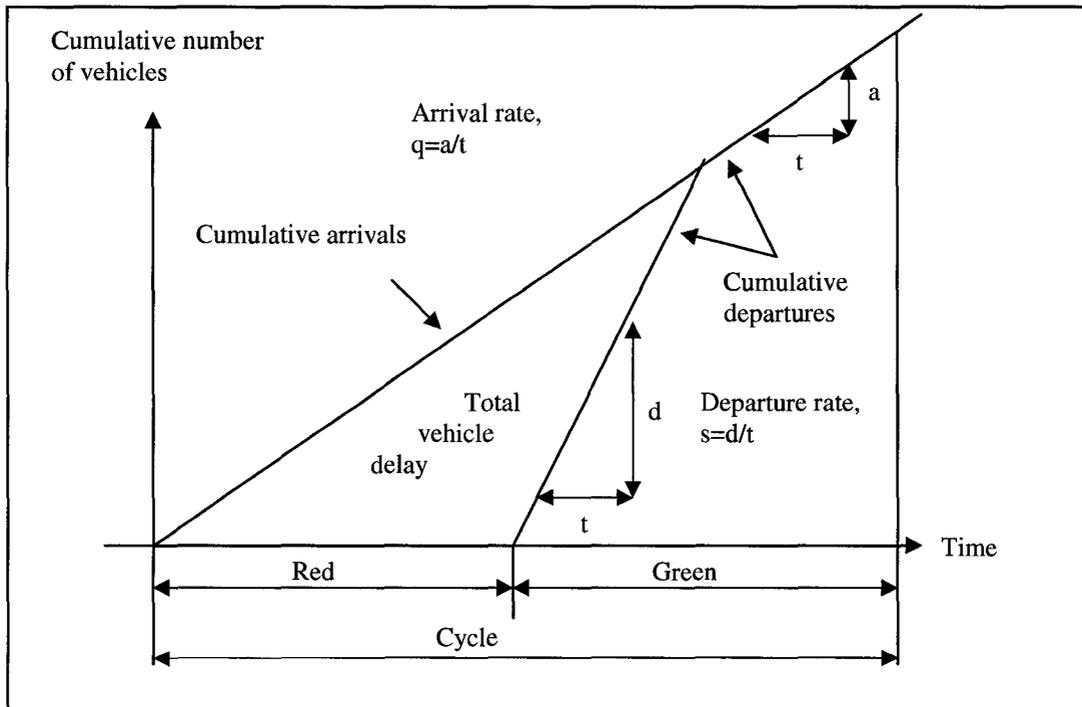


Figure 5.7: Uniform Delay

The second component d_2 (5.11) represents random or overflow delay. As the volume approaches capacity (but V/C is still less than one) due to the randomness of arrivals some of the vehicular queue overflows to the next cycle. The formula for d_2 has been derived theoretically (Hurdle, 1984, Allsop, 1972).

The third component d_3 , (5.12) is a reduction factor derived empirically through simulation (Webster, 1958). This term reduces the estimate by 5 to 15 percent (Hurdle, 1984).

By observing the second component d_2 , it is obvious that as the volume to capacity ratio (V/C) approaches 1, the second component approaches infinity. Conceptually, this is correct, if an intersection approaches its capacity, the queue grows continuously and delay increases indefinitely. Webster's steady state delay model assumes that the system operates for a sufficient time to settle and the settling time and the resulting delay is infinite.

However, in practice queues never grow infinitely, the time period when demand is higher than capacity is limited, and queues always start to dissipate after a certain period of time. Thus, the steady state assumption of the Webster's model does not hold, and the formula is not applicable at V/C ratios that approach unity, nor is it applicable to situations where the oversaturated conditions last for a limited time period.

For oversaturated conditions the deterministic queuing model is often used because it is potentially more accurate than the models based on Webster's original formulation such

as the CCG or HCM formulas. The graphical representation of the deterministic queuing model is shown in Figure 5.8. The cumulative departure curve is defined by the signal cycles (red + green interval). The horizontal part representing cumulative departures during the red signal interval (which is zero), and the sloped parts represent cumulative departures during the green interval. As shown previously in Figure 5.7, the slope represents the saturation flow (departure rate). The area between the cumulative departure and cumulative arrival curves represent the total vehicle delay. The slope of the dashed line represents capacity.

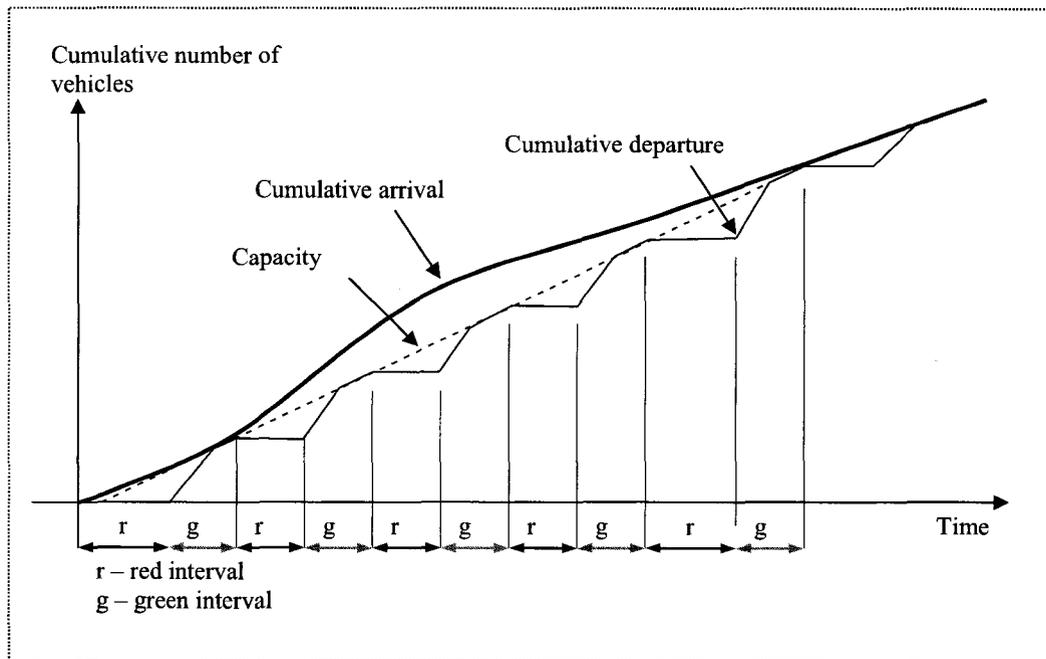


Figure 5.8: Graphical Representation of Oversaturated Conditions

The portion of area between the cumulative departure curve and the dashed line is described by the first component of Webster's delay formula 5.10. The portion of area between the dashed line and the cumulative arrival curve represents overflow delay.

Obviously, the area depends on the shape of the arrival curve, and to develop a generic delay model for oversaturated conditions, a simplified cumulative arrival curve was used as depicted in Figure 5.9 (Hurdle, 1984, Fambro, 1997).

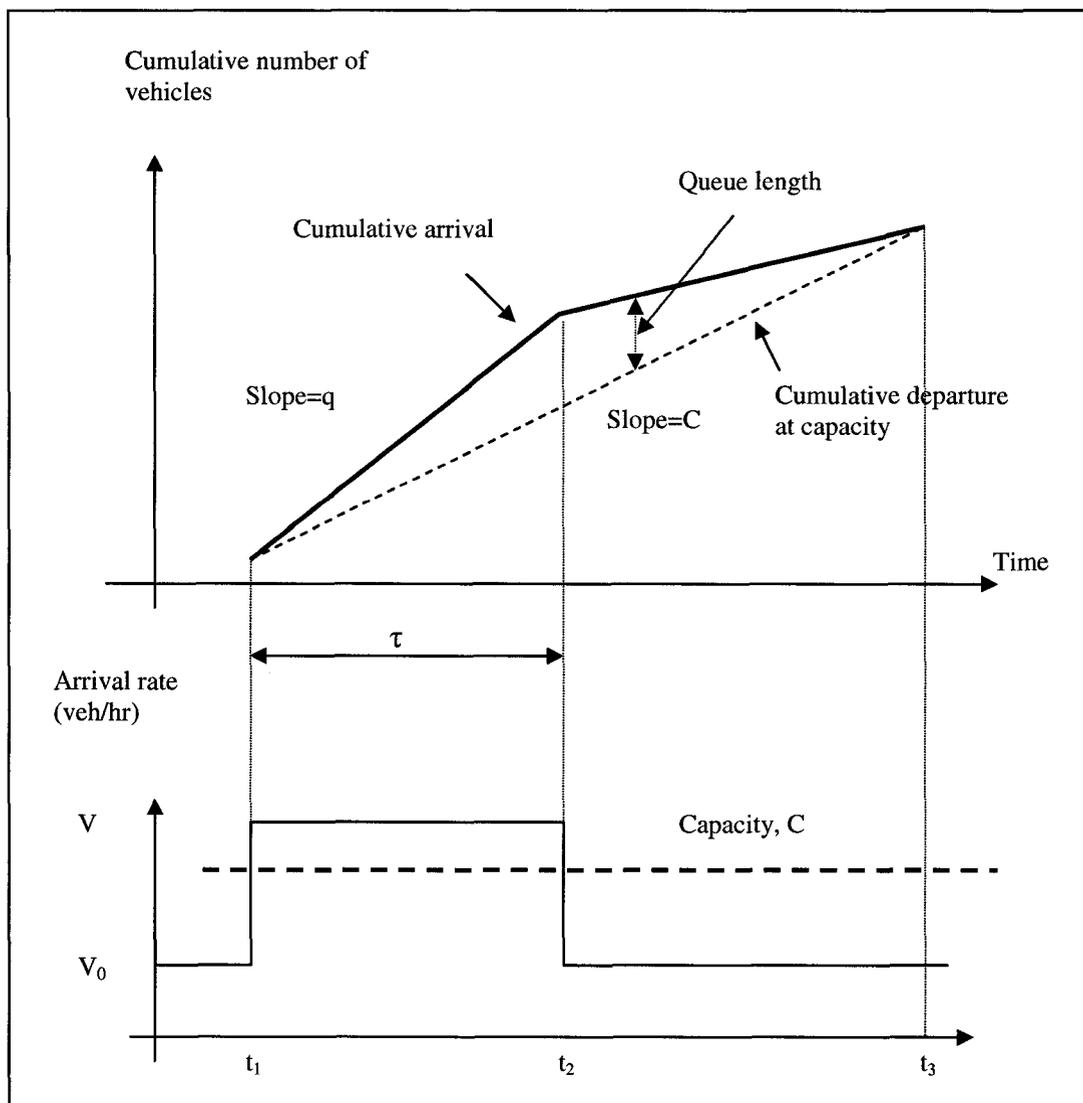


Figure 5.9: Simplified Representation of the Overflow Delay

The diagram on the bottom (Figure 5.9) shows a simple arrival pattern, the arrival rate jumps from V_0 to V at time t_1 and drops back to V_0 at time t_2 . Volume V is higher than

capacity, thus the cumulative arrival curve increases faster than the cumulative departure curve (top diagram). The queue reaches its maximum at t_2 and diminishes at t_3 and it is represented by a vertical line connecting the cumulative arrival and cumulative departure lines. Mean overflow delay can be calculated by dividing the area between the cumulative arrival and cumulative departure curves by the total number of vehicles affected, which is expressed by 5.13 (Hurdle, 1984).

$$d_2^{DQ} = \frac{\tau}{2} \times (X - 1) \quad 5.13$$

d_2^{DQ} - overflow delay based on the deterministic queuing model (sec/veh)

τ - time duration when the arrival flow is higher than capacity (sec)

X - degree of saturation

The total mean vehicle delay can be calculated by 5.14, where the first component represents Webster's uniform delay expressed by 5.10.

$$d = d_1 + d_2 \quad 5.14$$

Equation 5.13 reveals that the overflow delay, based on the deterministic queuing model increases linearly with the degree of saturation X , and the time duration while the arrival flow rate is higher than capacity.

The deterministic queuing model compliments Webster's model. While Webster's model covers the V/C region below unity quite reasonably, the deterministic queuing model appropriately covers the V/C region above unity. However, as Figure 5.10 reveals, neither of the two models provide reasonable values near unity. Webster's model, being a

steady state model, generates infinite overflow delays as the V/C approaches one, while the deterministic queuing model generates zero overflow delay at V/C equal to one. The queuing model generates zero overflow delay at $V/C=1$ because the model does not have a stochastic component thus ignores randomness which generates cycle overflows and significant additional delay.

The absence of the stochastic component from the deterministic queuing model at degrees of saturation significantly higher than one is less of a problem, because as the degree of saturation increases the impacts of randomness become relatively much smaller than the impacts of the constant overflow.

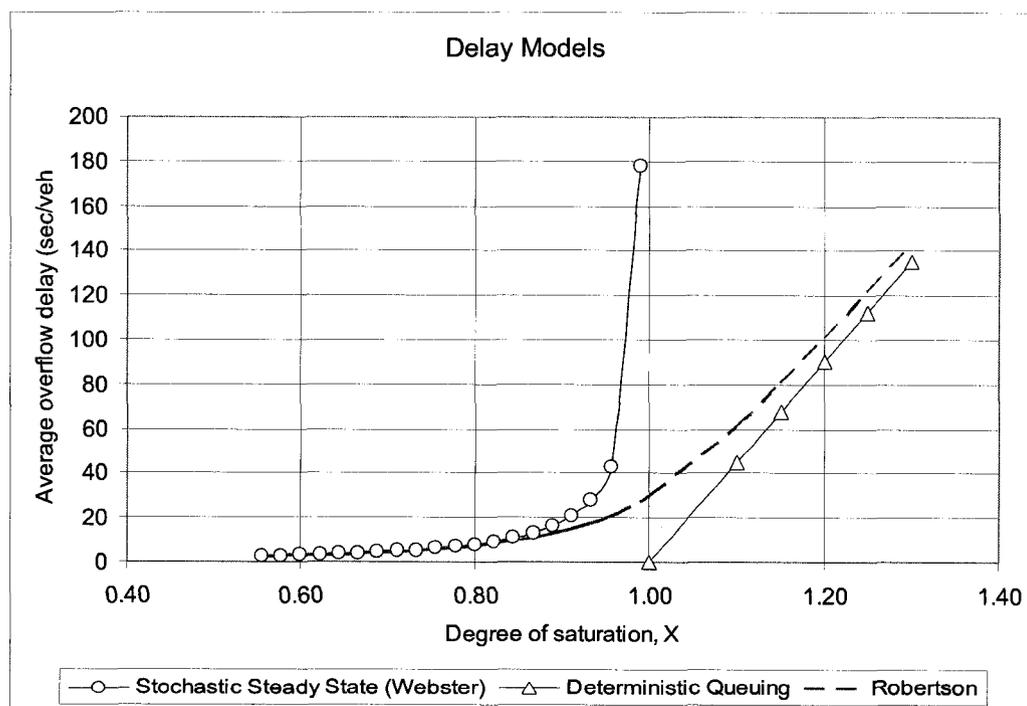


Figure 5.10: The Development of Oversaturated Delay Curve (cyc=60 sec, $g=30$ sec, $s=1800$ veh/hr, $t_a=15$ min)

To resolve the problem of incompatibility of the two models near unity, a coordinate transformation algorithm was proposed by P.D. Whiting (unpublished) who implemented the method in the Transport and Road Research Laboratory's TRANSYT model (Hurdle, 1984). The objective of the coordinate transformation was to modify the steady state curve to approach the deterministic oversaturation delay line asymptotically. A formula (5.15), approximating Whiting's method was published by Robertson (1979). The formula is time dependent equation since, as discussed previously, the overflow delay is influenced by the duration of time when the arrival flow is higher than capacity.

$$d_2 = 15 \times \frac{t_a}{C} \times \left[(V - C) + \sqrt{(V - C)^2 + 240 \times \frac{V}{t_a}} \right] \quad 5.15$$

d_2 – oversaturation delay (sec/veh)

t_a – duration of the analysis period (min)

C – capacity (veh/hr)

V – arrival volume (veh/hr)

The new oversaturation delay curve is also depicted in Figure 5.10 in a dashed line. It is important to re-iterate, that the curve is not the result of queuing analysis, it is a mathematical model of a transition curve from the steady state model to the deterministic queuing model.

The oversaturation delay expressed in Figure 5.10 has been adapted without or with some minor modifications in the Canadian Capacity Guide, the Australian, and the HCM delay models.

For example, in the Canadian delay model (CITE, 2008) oversaturated delay presented by 5.16 is exactly the same as Robertson's model expressed in a slightly different form.

$$d_2 = 15 \times t_a \times \left[(X - 1) + \sqrt{(X - 1)^2 + \frac{240 \times X}{C \times t_a}} \right] \quad 5.16$$

d_2 – mean overflow delay (sec/veh)

t_a – duration of the analysis period (min)

X – degree of saturation; $x=V/C$

V – volume (veh/hr)

C - capacity (veh/hr); $C=(g/cyc)*s$

cyc – cycle length (sec)

g - green time (sec/cycle)

s - saturation flow (veh/hr)

The random component in the Australian formula is defined by Equation 5.17 and Equation 5.18.

$$d_2 = \frac{t_a}{4} \times \left[(X - 1) + \sqrt{(X - 1)^2 + \frac{12 \times (X - X_0)}{C \times t_a}} \right] \quad 5.17$$

$$X_0 = 0.67 + \frac{s \times g}{600} \quad 5.18$$

When $X > X_0$, otherwise $d_2=0$

d_2 – mean overflow delay (sec/veh)

t_a – duration of the analysis period (sec)

X – degree of saturation; $X=V/C$

V – volume (veh/sec!)

C - capacity (veh/sec!); $C=g/cyc*s$

cyc – cycle length (sec)

g - green time (sec/cycle)

s' - saturation flow (veh/sec)

It is important to note that in the Australian model any units can be used but they must be consistent throughout the formula.

Finally, the random component in the HCM 2000 formula is defined by Equation 5.19 (TRB, 2000). The model has some additional parameters to capture the impact of actuated control and upstream metering. For the basic scenario with fixed time control and isolate operation, the formula provides identical results with the CCG.

$$d_2 = 900 \times t_a \times \left[(X - 1) + \sqrt{(X - 1)^2 + \frac{8 \times k \times l \times X}{C \times t_a}} \right] \quad 5.19$$

d_2 – incremental delay to account for effect of random and oversaturation queues adjusted for duration of analysis period and type of signal control (sec/veh)

t_a – duration of the analysis period (hr)

k – incremental delay adjustment for actuated control (fixed time control $k=0.5$)

l – incremental delay adjustment for filtering or metering by upstream signals (for isolated intersections $l=1$)

X – degree of saturation; $X=V/C$

V – volume (veh/hr)

C - capacity (veh/hr); $C=(g/cyc)*s$

cyc – cycle length (sec)

g - green time (sec/cycle)

s- saturation flow (veh/hr)

By comparing the Robertson, the CCG's, the Australian, and the HCM formulas, it is obvious that they are almost identical. However, compared to their predecessor, the deterministic queuing model, they differ in terms of the definition of the time dependent factor.

In case of the original deterministic queuing model, the definition of the time component ' τ ' is transparent, it represents the duration of time, when the arrival flow is higher than the capacity (Figure 5.9). The time component ' τ ' does not represent the period of congestion but the period when the queue increases. At the end of the period the queue is the longest but starts to dissipate. To calculate the mean vehicle delay from the moment t_1 when the queue starts to build up to moment t_3 when the queue has dissipated two parameters are needed, ' τ ' and V for that period. For the simple arrival pattern described in Figure 5.9, both parameters can be determined easily. Most importantly, time ' τ ' determines only the oversaturation delay, and has nothing to do with random delay.

In case of the four derived delay models, the definition of the time component is much less straightforward. Instead of representing the time when the arrival flow rate is higher

than capacity, the time component is defined as the *analysis period*, t_a . The difference between the two is not only semantics, it is conceptual. The analysis period ' t_a ' influences both the random and overflow delays because both are time dependent. However, the selection of t_a is unclear. Suppose that during a one hour period the V/C ratio never goes beyond 0.9. From the oversaturated conditions perspective, as defined in Figure 5.9, there is no need to calculate the overflow delay and the analysis period should be zero since the arrival flow rate is never higher than the capacity flow rate. But from the random delay perspective the analysis period is important since occasional queue overflows due to randomness are proportional to the time when volumes are close to capacity.

Various Traffic Engineering manuals deal with this problem differently. For example, the CCG suggests various values based on the population or other factors (Table 5.3). But no rationale has been provided for these values in the manual.

Table 5.3: Suggested Analysis Periods by the Canadian Capacity Guide (1995)

Regional population or other description	Analysis period or evaluation time, t_a (min)
<100 000	15 to 30
100 000 to 500 000	30 to 60
> 500 000	60
Special events	actual duration
Shift change in industrial areas (regardless of regional population)	15 to 30

The HCM 2000 default value for the analysis period t_a is 0.25 hr but the manual also suggests three alternative analysis approaches which is discussed in the following section.

5.3.1.2 The Impact of the Arrival Profile on Mean Vehicle Delay

Although the arrival profile has a significant impact on mean delay, none of the previously discussed delay formulas incorporate the pattern of queue build up and dissipation. The same number of arrivals, but distributed differently within the hour, would generate exactly the same mean vehicle delay when the delay formulas are used. This should not be surprising since the delay formulas are based on parameters describing signal operation and the mean volume but arrival profile information is not used at all.

The impact of the arrival profile on the mean delay is demonstrated on the following examples. Three different arrival profiles, as described on Figure 5.11, are examined.

(Other relevant input data are summarized in Table 5.4.)

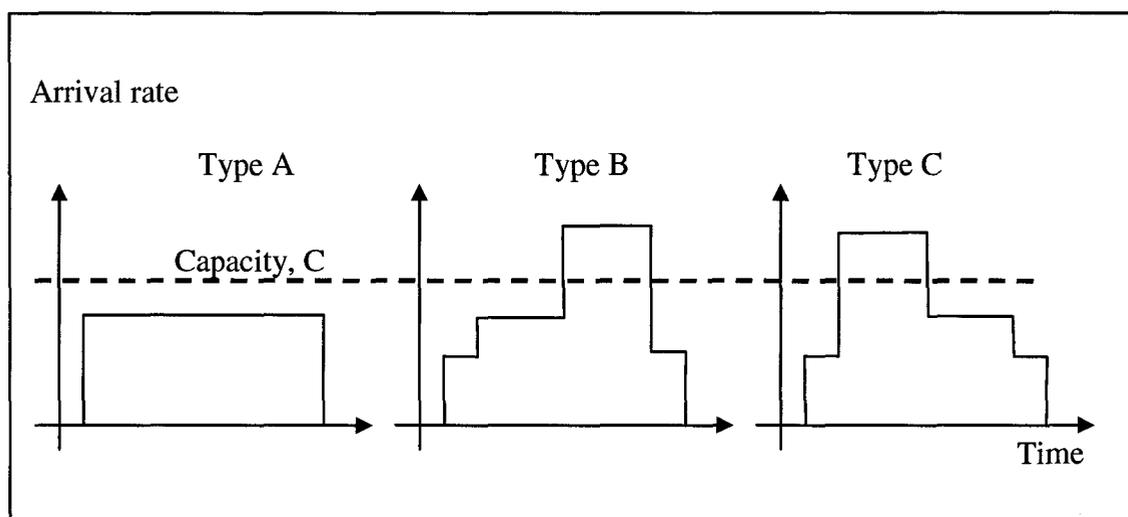


Figure 5.11: Description of Arrival Distributions A, B, and C

The capacity of the approach is 900 veh/hr ($30/60 \cdot 1800$) and demand is 700 veh/hr for all three arrival profiles. On average, the approach is unsaturated during the observation period with the same V/C ratio or D/C (demand over capacity) of 0.78 for all three cases.

In case of distribution A, the arrival rate is constant and below capacity. Thus there are no overflow queues and the D/C ratio is equivalent to the V/C ratio during the entire period of observation. In case of distributions B and C, for a certain period of time the D/C ratio is higher than the capacity thus there is queue build up and overflow queue from cycle to cycle.

Table 5.4: Input and Output Values for the Deterministic Queuing Model

Input Parameters	Values
Demand (veh/hr)	700
Cycle (sec)	60
Green (sec)	30
Saturation Flow (veh/hr)	1800
Demand/Capacity	0.78
Peak hour factor	0.662

A graphical presentation of the arrival distribution B and the cumulative arrival/departure curves is presented in Figure 5.12 and Figure 5.13. The curves were generated by an Excel spreadsheet programmed to calculate delay through the deterministic queuing model. (Further input/output details are provided in Appendix A)

The horizontal line at 75 vehicles/5 min (Figure 5.12) represents capacity. The stepwise line represents the distribution of arrival rate. The “zigzag” line represents the queue length with numerical values on the secondary y-axis. The “zigzags” are the result of queue build up and dissipation during cycles. Up to approximately 1800 sec, the queue clears in every cycle but starting from approximately 1800 sec the arrival rate increases

over capacity and the queue starts to build up until the arrival flow rate drops back to below capacity at 3000 sec. The built up queue dissipates at approximately 3300 sec.

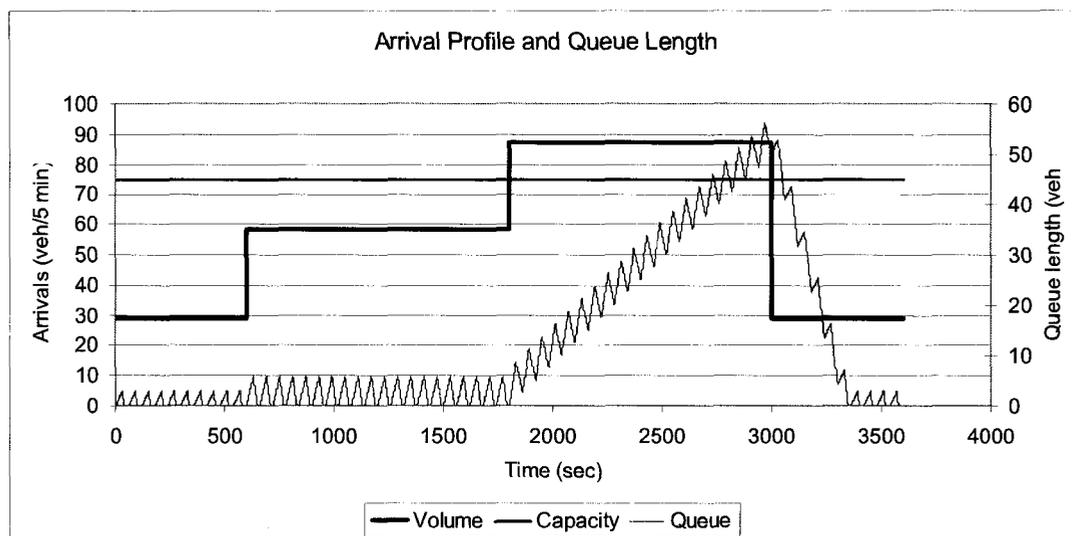


Figure 5.12: Arrival Distribution B

The cumulative arrival and departure curves (Figure 5.13) were constructed based on the arrival distribution curve. The vertical space between the cumulative arrival and departure curves represents the length of the queue, and the area between the two curves represents total vehicle delay. Up to approximately 1800 seconds, while the arrival flow rate is below capacity, total vehicle delay (secondary y-axis) includes only uniform signal delay and it grows slowly. After 1800 sec, when the arrival flow rate becomes larger than capacity, total vehicle delay increases rapidly due to the overflow delay. The growth in total vehicle delay drops when the built up queue dissipates.

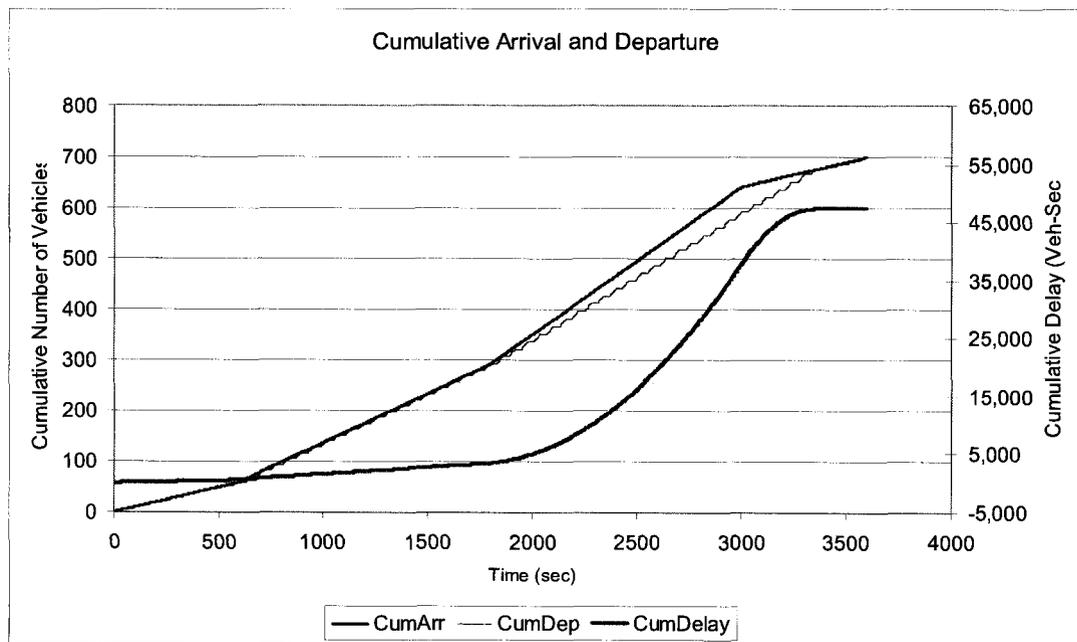


Figure 5.13: Cumulative Arrivals, Departures, and Delay for Arrival Distribution B

The arrival profile and cumulative arrival and departure curves for distribution C are presented in Figure 5.14 and Figure 5.15. By comparing the deterministic queuing models for distributions B and C, the difference between the cumulative arrival and departure curves, representing total vehicle delay, is obvious. Distribution B generates lower total vehicle delay.

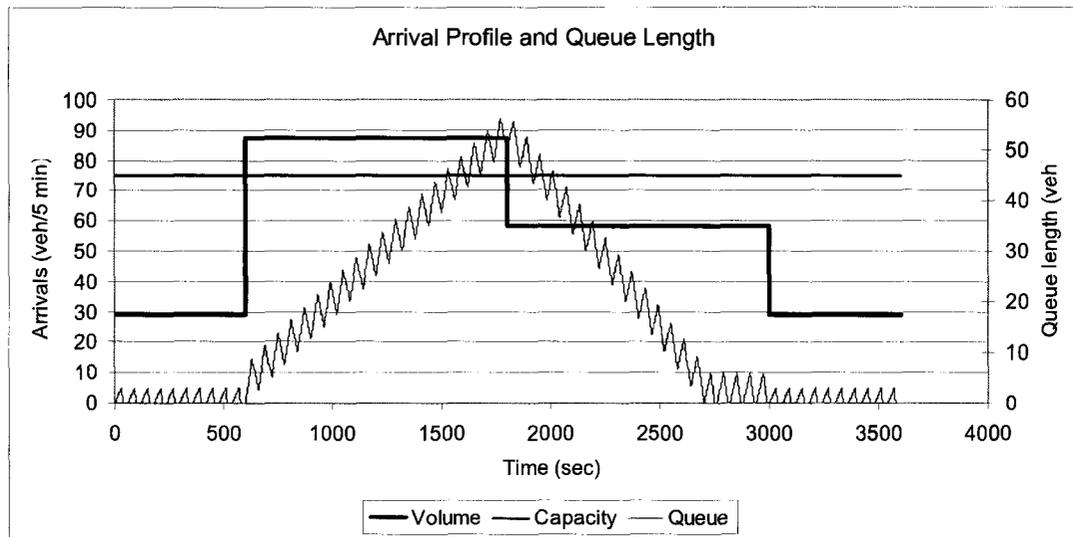


Figure 5.14: Arrival Distribution C

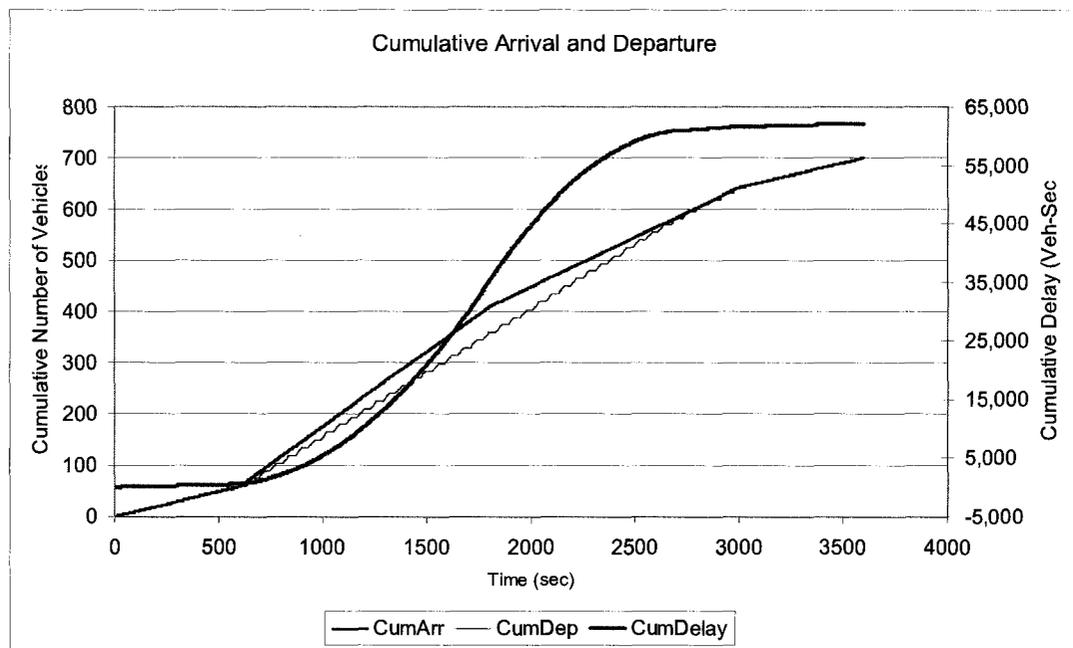


Figure 5.15: Cumulative Arrivals, Departures, and Delay for Arrival Distribution C

The analytical delay formulas do not use any information describing the arrival distribution, therefore, all three profiles generate the same delay values. The results of

delay calculations are summarized in Table 5.5. Delay calculation with the Webster, CCG, and HCM formulas are provided in (Appendix B).

The significant range of values derived by the various methods and assumptions is striking. The deterministic queuing model results are sensitive to different arrival profiles, and the difference between uniform arrivals (distribution A) and the non-uniform arrivals (distributions B and C) are quite significant.

Table 5.5: Mean Delay Calculated by the Deterministic Queuing Model and Analytical Formulas

Mean vehicle delay (sec/veh)					
Deterministic queuing model			Webster	CCG	HCM
Distribution A	Distribution B	Distribution C	Distributions A,B,C		
12	68	89	17.4	18.8	18.8

The Webster, the HCM, and the CCG delay formulas provide somewhat higher values than the deterministic queuing model for the uniform arrival (distribution A) because in addition to the uniform delay captured by the deterministic queuing model they also include a random delay component. However, the formulas significantly underestimate vehicle delay for non-uniform arrivals described by distributions B and C. This is a serious shortcoming of the formulas since oversaturated conditions have always non-uniform arrival distribution. The arrival rate always starts from a value that is lower than capacity, it increases above capacity, and finally it drops below capacity. Although this process can be expressed by the mean arrival rate, the delay calculation based on this number is extremely unreliable.

The HCM 2000 recognized the shortcoming of the delay formula for assessing oversaturated conditions and suggests three different analysis approaches (Figure 5.16).

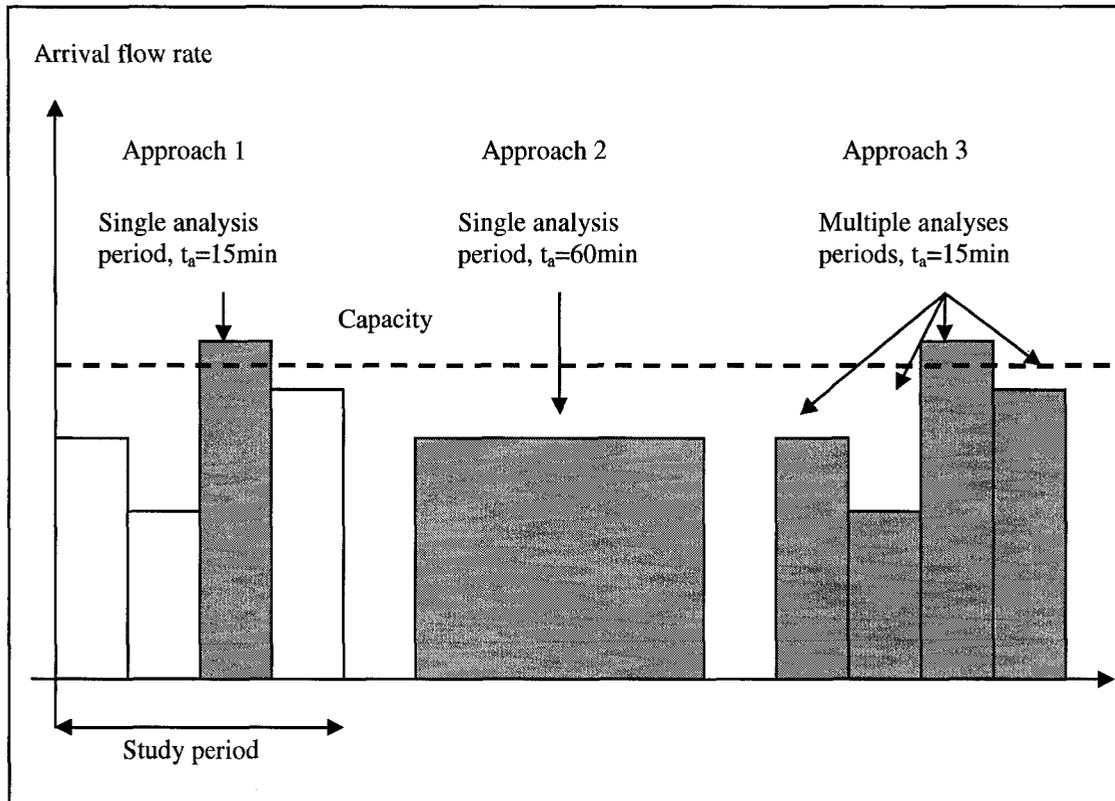


Figure 5.16: Alternative HCM 2000 Approaches for the Analysis of Oversaturated Conditions (Source: TRB, 2000)

Approach 1 is the traditional HCM approach where the $t_a = 15$ min peak period is used for the analysis. The underlying objective is to design signal operation for the worst 15 min period within the peak hour, thus the peak 15 min period volume is used. This concept, however, breaks down when the demand flow during the peak 15 min period is above capacity. In that case, the traditional HCM approach significantly overestimates mean

delay because it assumes an arrival rate that is higher than capacity. (Calculation is provided in Appendix B).

Approach 2 takes the mean demand over the study period and uses $t_a=60$ min as the analysis period. This approach has two risks. The first risk is the underestimation of mean delay because by taking the mean demand over the analysis period, the impacts of peaks in the arrival flow are ignored. The second risk is overestimating mean vehicle delay because of the long analysis period, t_a . The potential for risk one and two are determined by the demand (arrival) distribution; more uniform arrivals would increase overestimation and higher peaking would increase underestimation of the mean delay.

Approach 3 divides the study period into four $t_a=15$ minute analysis periods and the 15 min demand rates are used. The impact of overflowing queues during analysis periods when demand is over capacity is also included in the delay calculation. (Calculation is provided in Appendix B.)

The mean vehicle delay calculated by Approach 1 and Approach 2 are compared to the previous HCM delay calculation. An attempt has been made to use Approach 3 as well, however, the HCM description of the procedure is not complete. Table 5.6 shows the input volumes including the volume and the analysis period, t_a , and the results of the delay calculations. (Calculations are provided in Appendix B.)

Table 5.6: Results of Vehicle Delay Calculation in Oversaturated Conditions Based on Different Procedures Presented in HCM (2000)

	HCM	HCMA_1	HCMA_2	HCMA_3
Analysis volume, V (veh/hr)	700	700/0.662 =1057	700	different for each analysis period
Period of analysis, t _a (hr)	0.25 hr	0.25 hr	1 hr	duration of individual analysis periods
Mean delay, d (sec/veh)	18.8	105	19.2	N/A ¹⁾

¹⁾ The HCM 2000 procedure is not complete (page 16-148 TRB, 2000)

Based on the above analysis, it seems obvious that neither Approach 1 nor Approach 2 is accurate. Furthermore, a theoretical foundation for determining the appropriate analysis period, t_a is also missing from the HCM delay calculation description.

Both the CCG and the HCM delay formulas include an adjustment factor for progression that modifies the uniform delay component based on the arrival profile within the cycle. However, the arrival profile within the cycle should not be confused with the arrival profile within the period of observation; the first is influenced by signal coordination, and the second is determined by the dynamics of demand increase and dissipation.

Based on the above analysis it can be concluded that the traditional analytical formulas presented above are unreliable and inadequate for estimating vehicle delay in oversaturated conditions.

5.3.2 Transit Modal Share (TMS)

The increase in transit modal share is a fundamental objective of bimodal traffic management in oversaturated conditions. However, predicting TMS changes due to various traffic control/management strategies is not a simple process.

The estimation of mode share has been introduced to transportation planning as one element of the Urban Transportation Modeling System (UTMS) originally developed in the 1960s for macro scale regional planning (Weiner, 1992). The objective of UTMS is to predict the number of trips by purpose and mode during various time periods on individual links or between origin-destination pairs and it was developed primarily to support highway planning.

The third stage in the UTMS, the mode choice (Figure 5.17), is the calculation of the mode split or in our case the TMS. TMS describes the role of transit in satisfying the demand for travel and it is expressed as the ratio of people travelling by transit over the total number of people (travelling by car and transit vehicles) (5.20).

$$TMS = \frac{NT_T}{NT_T + NT_C} \quad 5.20$$

TMS – transit modal share

NT_T – number of travellers on transit vehicles

NT_C - number of travellers in cars

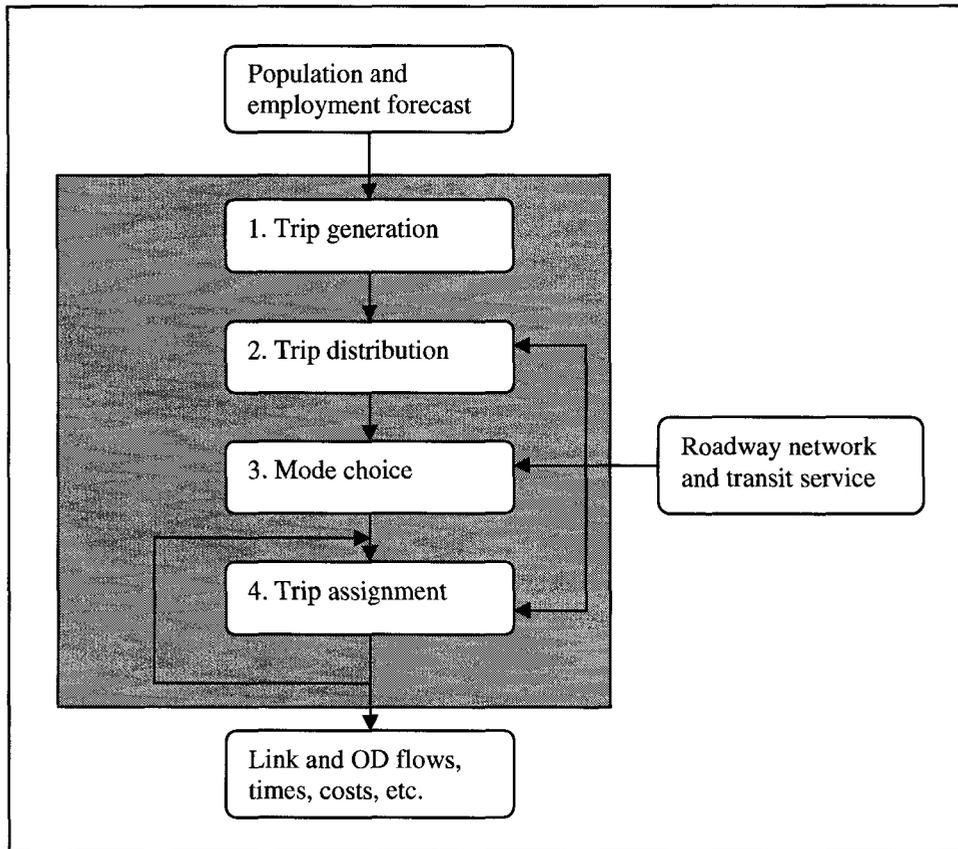


Figure 5.17: The Urban Transportation Modeling System (UTMS)

The transportation planning literature contains a number of mode choice models (Meyer et al., 2001, Ben-Akiva et al., 1985) and most mode choice models are based on the logit formulation. The general form of the multinomial logit model is given by 5.21.

$$P_i = \frac{e^{u_i}}{\sum_{i=1}^j e^{u_i}} \quad 5.21$$

P_i – probability that a traveller will chose mode i for the trip

u_i – systematic (observable) utility of mode i

j – the number of alternative modes of travel

The logit model estimates the probability of an individual choosing a specific mode of transportation based on the relative utility of that mode compared to all the other alternatives. The utility of a particular mode has the following general form (5.22) (NCHRP, 1998).

$$u_i = a_i + b_i IVTT_i + c_i OVTT_i + d_i COST_i \quad 5.22$$

u_i – utility of mode i

a_i – mode bias coefficient for mode i

b_i, c_i, d_i – coefficients for mode i for the $IVTT_i, OVTT_i,$ and $COST_i$ respectively

$IVTT_i$ - service attribute in-vehicle travel time for mode i (min)

$OVTT_i$ – service attribute out-of-vehicle travel time for mode i (min)

$COST_i$ – service attribute cost of travel by mode i (\$)

The utility of a mode describes the attractiveness of the mode based on coefficients (e.g. a_i, b_i, c_i, d_i) and level of service attributes (e.g. $IVTT, OVTT,$ and $COST$). Coefficient a_i is a mode specific coefficient and it describes the general preference of the population for the mode. Service attributes could be either generic or mode specific. For example some models treat $IVTT$ as a mode-specific attribute based on the assumption that a specific time spent on a transit vehicle is perceived differently than the same amount of time spent in a car. The $OVTT$ and the $COST$ are often assumed to be independent of the mode. Whether a service attribute is generic or mode-specific is important because it impacts the number of different coefficients that have to be calibrated. Generic service attributes

have one coefficient for all modes and mode specific attributes need separate coefficient for each mode.

The utility equation 5.22 has a relatively simple form. But mode attributes IVTT, OVTT, and COST are often further broken down to subcomponents. Table 5.7 shows the main attributes, their potential subcomponents and the range of coefficients based on data from 13 US urban areas by Parsons(1992) and KPMG (1992).

Table 5.7: Mode Choice Attributes and their Coefficients (Parsons et al., 1992; KPMG, 1992; NCHRP 1998)

Main mode attributes	Mode attributes and/or subcomponents	Coefficients
IVTT	In vehicle travel time	-0.015 to -0.04
	In vehicle travel time to transit terminal	-0.03 to -0.286
OVTT	Out of vehicle time	-0.028 to -0.112
	Walk time to transit	-0.033 to -0.114
	Wait time for transit	-0.023 to -0.77
	Transit transfer time	-0.032 to -0.114
COST	Total cost	-0.0037 to -0.014
	Transit fare	-0.004 to -0.005
	Car operating cost	-0.004 to -0.005
	Parking cost	-0.009 to -0.012

It can be observed that although range of coefficients is significant they have roughly the same order of magnitude for the 13 US cities. The calibration of utility models is data intensive and the final model is location specific. Furthermore, the model presented in 5.21 applies to a specific individual traveller and to estimate transit modal share for an entire corridor or urban area, the mode choices of individuals have to be aggregated for a particular zone or study area (Meyer, 2001).

To facilitate the estimation of impacts of various measures on transit modal share, NCHRP Report 365 (1998) suggests the application of an *incremental logit model*. The advantage of the incremental logit model is that instead of describing the relationship between all independent variables (modal attributes) and the probability of choosing a specific mode (or modal share), it uses existing probabilities and only those service attributes that are impacted by the modification. The incremental logit model is presented in Equations 5.23 where Δu_i represents the change in the utility of the mode due to the modification (5.24) (NCHRP, 1998).

$$P_i' = \frac{P_i \times e^{\Delta u_i}}{\sum_j (P_j \times e^{\Delta u_j})} \quad 5.23$$

$$\begin{aligned} \Delta u_i &= (a_i + b_i \times IVTT_i' + c_i \times OVTT_i' + d_i \times COST_i') - \\ &\quad (a_i + b_i \times IVTT_i + c_i \times OVTT_i + d_i \times COST_i) = \\ &= b_i \times \Delta IVTT_i + c_i \times \Delta OVTT_i + d_i \times \Delta COST_i \end{aligned} \quad 5.24$$

P_i – existing probability of traveller choosing mode i (or measured mode share)

Δu_i – change in utility for mode i

$IVTT_i'$, $OVTT_i'$, $COST_i'$ – service attribute before the change

$IVTT_i$, $OVTT_i$, $COST_i$ – service attribute after the change

$\Delta IVTT_i$, $\Delta OVTT_i$, $\Delta COST_i$ – change in the service attribute value

Although its application is limited to problems where certain baseline modal share data already exists, the incremental logit model has three significant advantages compared to its parent model.

First, the mode specific parameter a_i is dropped from the utility function (5.24). The impact of the preference for the mode is already captured by P_i - the existing probability for selecting the mode.

Second, service attributes that do not change are also eliminated from the utility function. For example if the implemented traffic engineering measures impacts only bus IVTT (in vehicle travel time) and all other service attributes for all modes remain constant, the utility functions for bus and car modes can be expressed by 5.25 and 5.26 respectively.

$$\Delta u_{bus} = b_{bus} \times \Delta IVTT_{bus} \quad 5.25$$

$$\Delta u_{car} = 0 \quad 5.26$$

Third, the incremental logit model is more transferable between urban areas than the simple logit model because the impacts of service attributes on the probability of selecting a specific mode has less variation between different urban areas than the mode bias coefficient a_i which captures factors that are difficult to define and measure (NCHRP, 1998).

Based on the NCHRP (1998) report, the range of coefficients b_i and c_i , calibrated in 13 US urban areas, is not too large thus mean values are suggested in the absence of site specific calibration data. The suggested mean value for b_i and c_i are -0.025 and -0.050, respectively.

The application of the incremental logit model will be examined on the following example. Assume that the transit modal share between a specific origin and destination

pair is 10%. Due to the implementation of transit priority measures transit travel time decreased by 10 min. It is expected, that the relative improvement in transit travel times compared to car travel times would have a positive impact on transit ridership and modal share for the specific OD pair. The change in the utility of the bus and car modes is expressed by equations 5.27 and 5.28 and the resulting modal shares are calculated by using 5.29 and 5.30.

$$\Delta u_{bus} = b_{bus} \times \Delta IVTT_{bus} = -0.025 \times (-10) = 0.25 \quad 5.27$$

$$\Delta u_{car} = 0 \quad 5.28$$

$$P'_{bus} = \frac{P_{bus} \times e^{\Delta u_{bus}}}{P_{bus} \times e^{\Delta u_{bus}} + P_{car} \times e^{\Delta u_{car}}} = \frac{0.1 \times e^{0.25}}{0.1 \times e^{0.25} + 0.9 \times e^0} = 0.125 \quad 5.29$$

$$P'_{car} = \frac{P_{car} \times e^{\Delta u_{car}}}{P_{bus} \times e^{\Delta u_{bus}} + P_{car} \times e^{\Delta u_{car}}} = \frac{0.9 \times e^0}{0.1 \times e^{0.25} + 0.9 \times e^0} = 0.875 \quad 5.30$$

The relationship between TMS and the transit travel time change based on the incremental logit mode choice model is presented in Figure 5.18 with the line without markers. A 10 minute transit travel time improvement, while everything else remains unchanged, would result in an increase in transit modal share from the existing 10% to approximately 12.5%. The direction of change is logical and the order of magnitude of the change is within the limits of expectations, however, studies outside of this research would be helpful to confirm these values and their variability.

It is obvious that transit travel time reduction has a direct positive impact on transit service attractiveness. However, travel time savings may have impact on the out-of-

vehicle travel time (OVTT) as well. The magnitude of this impact depends on the decision of the transit agency regarding the utilization of the travel time savings. In principle, there are three alternatives to utilize the benefits of transit travel time savings:

- **To reduce the number of transit vehicles on the route:** as transit travel times decrease, at one point the reduction is sufficient to reduce the number of vehicles and provide the same service frequency and capacity (i.e. shorter travel time, same schedule at stops, savings used to reduce the number of buses)
- **To improve schedule reliability:** if mean transit travel time is reduced while the maximum allocated trip time remains the same, the improvement serves to enhance the reliability of the transit system (i.e. shorter travel time, same schedule at stops, savings used as “reserve/recovery” time)
- **To increase the frequency of the service:** as transit vehicles need less time to make the trip, they could start the next trip earlier, thus both the service frequency and capacity would increase (i.e. shorter travel time, same number of buses, savings used to increase frequency)

If the first and second approaches are used, the impacts on TMS could be calculated as previously demonstrated through the incremental logit model. However, if travel time savings are used to increase the frequency of service, the mean wait time for the service and the OVTT will decrease. Assuming that the mean wait time can be described as half of the service headway (time between two buses) the mean wait time can be expressed by

5.31 and the relationship between travel time saving $\Delta IVTT$ and mean wait time saving $t_{\Delta w}$ can be expressed by 5.32

$$t_w = \frac{1}{2} \times \frac{t_{RT}}{n} \quad 5.31$$

$$\begin{aligned} t_{\Delta w} &= \frac{1}{2} \times \frac{t_{RT} + \Delta IVTT}{n} - t_w & 5.32 \\ &= \frac{1}{2} \times \frac{t_{RT} + \Delta IVTT}{n} - \frac{1}{2} \times \frac{t_{RT}}{n} \\ &= \frac{\Delta IVTT}{2 \times n} \end{aligned}$$

t_w – mean wait time (min)

$t_{\Delta w}$ – change in mean wait time (min)

t_{RT} – transit round trip time (min)

n – number of buses

$\Delta IVTT$ – change in in-vehicle travel time (min)

Assuming that beside the reduced wait time nothing else changes in terms of the out-of-vehicle travel time (OVTT), the change in the OVTT is equal to the change in the mean wait time (5.33).

$$\Delta OVTT_{bus} = t_{\Delta w} = \frac{\Delta IVTT_{bus}}{2 \times n} \quad 5.33$$

Based on the above discussion, travel time reduction would have an impact on OVTT (out-of-vehicle travel times), thus the transit modal share calculation would have to include that impact as well and the new expression describing the change in the utility of transit can be described by 5.34.

$$\begin{aligned}
\Delta u_{bus} &= b_{bus} \times \Delta IVTT_{bus} + c_{bus} \times \Delta OVTT_{bus} & 5.34 \\
&= b_{bus} \times \Delta IVTT_{bus} + c_{bus} \times \frac{\Delta IVTT}{2 \times n} \\
&= \Delta IVTT_{bus} \times \left(b_{bus} + \frac{c_{bus}}{2 \times n} \right)
\end{aligned}$$

As Equation 5.32 indicates, beside the change of in-vehicle travel time, the number of buses servicing the stop has an impact as well. As the number of buses increases, the impact of travel time change on the change of wait time decreases. Assume that the existing TMS is 10 %. A 10 minute travel time reduction ($\Delta IVTT$) would increase TMS to approximately 13% if 4 buses were providing the service and to 15.5% if only one bus was providing the service (Figure 5.18) . The reason for this is simple, with more buses on the same route, the headway and the associated wait time (OVTT) is shorter, thus the impact of travel time change on the wait time is smaller.

Ideally, the evaluation of transit modal share is an iterative process. Relative improvements to the transit travel times compared to car travel times increase transit modal split by attracting existing car drivers. However, as the number of vehicles decrease on the network, vehicle travel times improve as well, which impacts the TMS calculation. To calculate the long-term impact of a traffic control/management approach, the system has to reach a steady state. To determine the steady state TMS, the iterative process, described in Figure 5.19, would have to be carried out. However, not all steps described in the algorithm can be calculated through formulas. As discussed previously, delay at congested intersections and consequently the travel time depends on the shape of the arrival distribution, thus there are no universal analytical expressions to describe it.

The functional relationships between intersection delay and the level of saturation developed for planning models assume average conditions and they were not designed for oversaturated conditions. Using travel times from simulation models might be a more accurate approach.

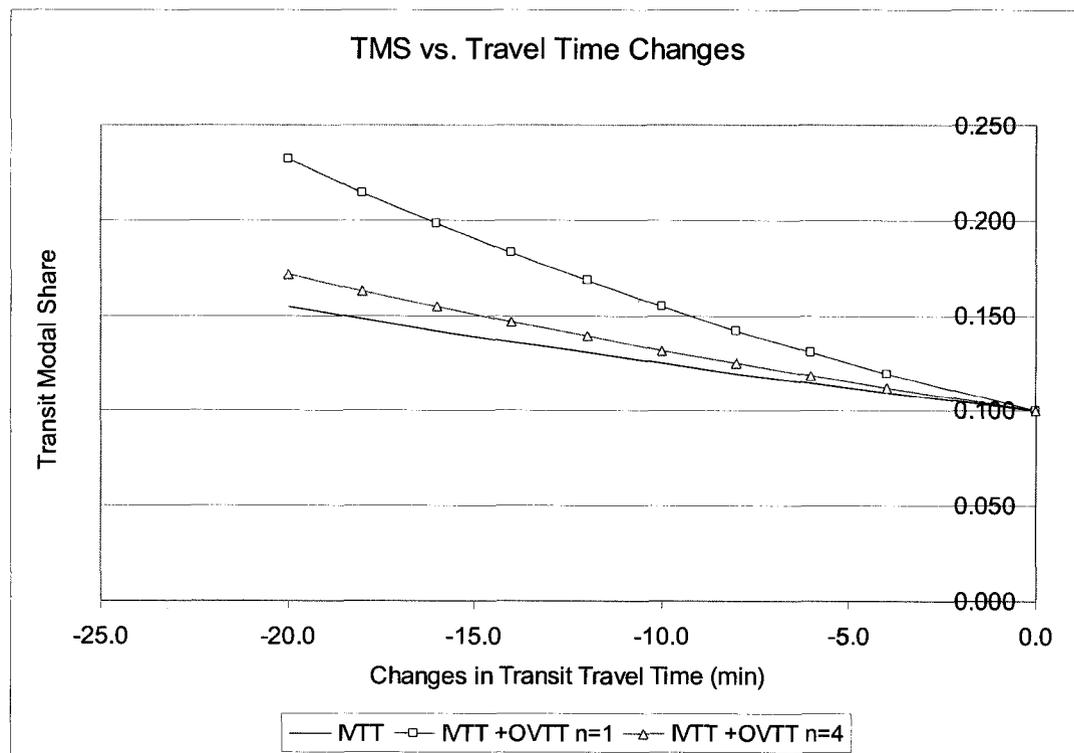


Figure 5.18: The Impact of Transit Travel Time Change on Transit Modal Share while Everything Else Remains the Same

Based on the above discussion it can be concluded that the calculation of TMS due to travel time changes is a relatively complex process thus expecting a routine implementation of TMS as a MoP in assessing Queue Management strategies may be

unrealistic. But, the impacts of bimodal queue management measures on TMS cannot be ignored since the primary reason for their implementation is to increase the TMS.

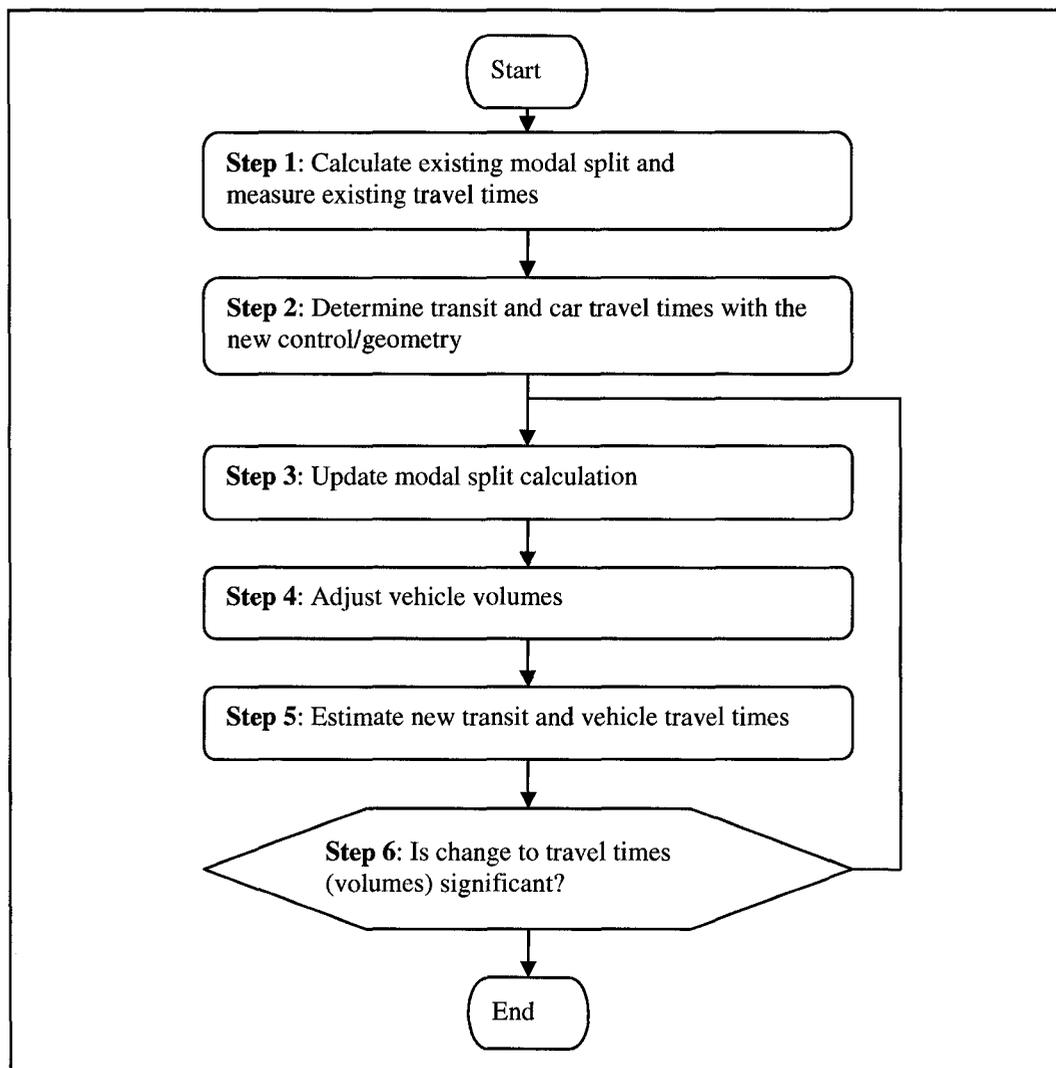


Figure 5.19: Iterative Modal Split Calculating Algorithms

Therefore, a simpler, more practical approach is needed to establish a connection between bimodal transportation measures and the TMS. The relationship between TMS and relative advantage of transit travel times compared to car travel times is explored next.

Assume that transit travel time savings through traffic management measures are all used to increase the frequency of the service. The change in the utility of the bus mode has been already defined in Equation 5.24. The expression for the change of utility function (Δu_{bus}) has to be transformed to have the following format (Equation 5.35):

$$\Delta u_{bus} = f(TCTTr) \quad 5.35$$

The transformation steps are described by the following equations:

$$\Delta IVTT_{bus} = IVTT_{bus}^{new} - IVTT_{bus}^{old} \quad 5.36$$

$$TCTTr^{old} = \frac{IVTT_{bus}^{old}}{IVTT_{car}^{old}} \quad 5.37$$

$$IVTT_{bus}^{old} = TCTTr^{old} \times IVTT_{car}^{old} \quad 5.38$$

$$TCTTr^{new} = \frac{IVTT_{bus}^{new}}{IVTT_{car}^{new}} \quad 5.39$$

$$IVTT_{bus}^{new} = TCTTr^{new} \times IVTT_{car}^{new} \quad 5.40$$

By inserting 5.38 and 5.40 into 5.36 we derive 5.41.

$$\Delta IVTT_{bus} = TCTTr^{new} \times IVTT_{car}^{new} - TCTTr^{old} \times IVTT_{car}^{old} \quad 5.41$$

Assuming car travel time does not change (Equation 5.42), we derive Equation 5.43.

$$IVTT_{car} = IVTT_{car}^{new} = IVTT_{car}^{old} \quad 5.42$$

$$\begin{aligned} \Delta IVTT_{bus} &= TCTTr^{new} \times IVTT_{car} - TCTTr^{old} \times IVTT_{car} \\ &= IVTT_{car} \times (TCTTr^{new} - TCTTr^{old}) \end{aligned} \quad 5.43$$

And finally, by inserting Equation 5.43 into Equation 5.34 we get an expression describing the change of transit utility as a function of the relative change of transit travel times compared to car travel times (Equation 5.44).

$$\Delta u_{bus} = IVTT_{car} \times (TCTTr^{new} - TCTTr^{old}) \times \left(b_{bus} + \frac{c_{bus}}{2 \times n} \right) \quad 5.44$$

By using Equation 5.44 to calculate the change on the utility of the transit mode and Equation 5.23 to calculate the new TMS, the relationship between the relative advantage of transit travel time (TCTTr) and TMS is established (Figure 5.20). As the TCTTr decreases (relative transit improvements) the TMS increases. The lower and upper limits were generated by using different service attribute coefficients b for IVTT and c for OVTT (Equation 5.14, and Table 5.7, Appendix D). In the absence of local modal calibration data, using a range of likely TMS outcomes is the best next alternative.

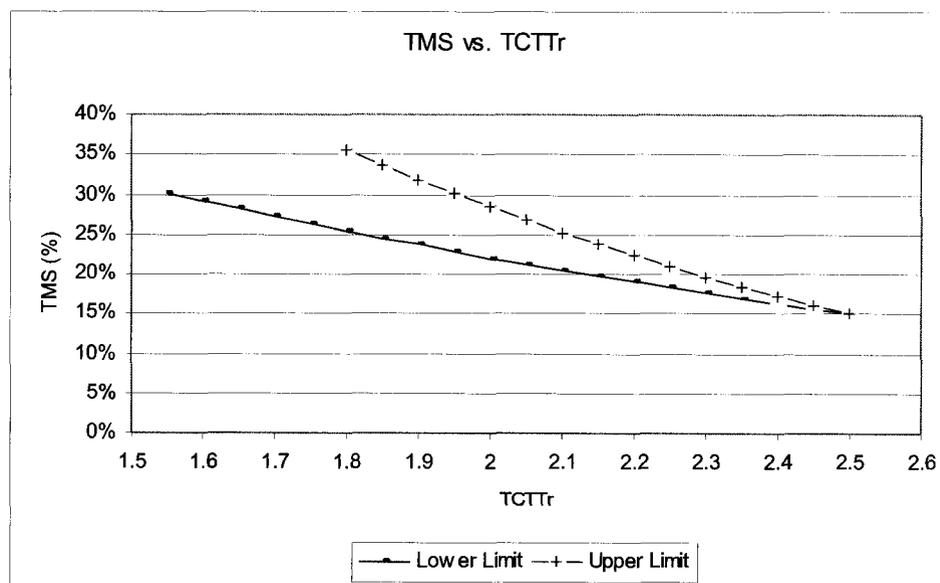


Figure 5.20: The General Relationship between TMS and TCTTr

Based on the TMS vs. TCTTr relationship, one can determine the design and operational requirements, from the roadway infrastructure perspective, that is required to achieve TMS policy objectives. For example, if the objective is to double TMS for a particular origin destination pair within 20 years from the existing 15%, modifications to the infrastructure over the 20 year period would have to ensure that the existing TCTTr of 2.5 decreases to a value between 1.55 and 1.95 approximately.

The TCTTr for mixed flow corridors of a 5 to 10 km length where transit travel times are not shielded from the impacts of congestion, is typically between the value of 2 and 3. For corridors where transit is not impacted by congestion, the TCTTr can easily reach values between 1 and 1.5 (Mucsi, 2004)

There are a number of other factors beside the TCTTr that influence transit ridership and TMS, however, the TCTTr is predominantly within the domain of the traffic engineering discipline, and is influenced by infrastructure design and day-to-day operational decisions.

One of the major challenges in using the TCTTr in predicting the change in TMS is the process of determining the relevant origin-destination (OD) pair. Calculating or measuring the TCTTr change for a short roadway section where the queue management techniques was implemented is simple. However, to measure the impact on TMS, the savings have to be spread out to the entire trip between the relevant OD pair.

For example, assume that transit travel between A and B before the measure is 10 min and car travel time is 8 min. The TCTTr is $10/8=1.25$. With bimodal queue management transit travel time decreases to 5 min, car travel time remains unchanged, thus the TCTTr is $5/8=0.625$. The relative improvement in transit travel times expressed by the change of the TCTTr (from 1.25 to 0.625) cannot be used to determine the impact on TMS, because points A and B (Figure 5.21) do not represent OD points to travellers.

Let's assume that the travel time, without the measure, between OD is 120 min and 60 min for transit and cars respectively. When the 5 min transit travel time saving is spread out to this trip, the TCTTr change is from 2 ($120/60=2$) to 1.9 ($(120-5)/60=1.9$). This change of TCTTr can be used to estimate the impact on TMS as explained earlier.

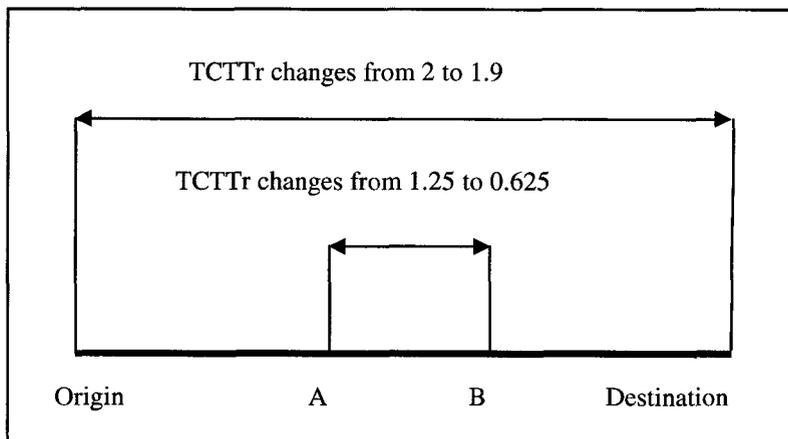


Figure 5.21: The Expansion of TCTTr from a Short Section to the OD Pair

Determining the relevant origin-destination points that are affected by a particular measure (and travel time saving) is possible through regional scale transportation planning models (e.g. EMME/2). Although that approach has merits theoretically, it is infeasible in most practical cases. First, not every urban area has a transportation

planning model. Second, the models are usually not calibrated with the required accuracy for modeling travel time changes of a few minutes. Third, the amount of modelling work required for the evaluation is way out of the scope of any evaluation budget. In comparison, the assessment of infrastructure or operations modifications from the single mode perspective is often done through a simple Volume over Capacity (V/C) calculation. Asking to run a regional planning model to justify a single bimodal queue management measure is considered excessive.

The challenge is great but not uncommon: It is known that a strong relationship between TMS and TCTTr exists, but there are no simple and practical ways to quantify the relationship. This leaves three alternatives:

- **Ignore the connection between TCTTr and TMS:** Currently, this is the prevailing approach. However, by pursuing this approach it is difficult to justify the implementation of bimodal queue management measures because the long-term benefits of queue management are ignored or underestimated.
- **Use the TCTTr calculated for the short section where the queue management was implemented:** Obviously, this approach greatly overestimates the benefits.
- **Select an origin destination pair based on some simplified logic:** This approach is between the first two, and is based on a somewhat arbitrary selection of origin-destination points.

The first alternative is clearly not palatable. The saying “what is not measured cannot be improved” applies. The second alternative is more useful. Although it cannot be used to

quantify the impacts on TMS, it can be used for comparison purposes in the evaluation of different queue management techniques. It can be also used as a qualitative tool to establish which infrastructure design and operations control measure has a positive, negative, or neutral impact on TMS. However, the third alternative has the most potential, but it needs further research and development work which is outside of the scope of this thesis.

5.4 Proposed Performance Measures

Although both examined MoP (e.g. vehicle delay and TMS) are frequently used in Transportation Engineering/Planning, their direct utilization in the evaluation of control in multi-modal oversaturated conditions is problematic. Existing analytical vehicle delay calculation procedures are inadequate for oversaturated conditions and existing TMS calculation procedures are too complex for practical day-to-day applications.

Furthermore, the TMS measure cannot be used directly to support design and operations decisions.

Bimodal MoP must satisfy the following criteria to be practical and effective:

- relatively simple to calculate;
- ability to capture the bimodal aspect of operation; and
- ability to connect transportation planning, system design, and system operation objectives.

The selection of MoP depends on the perspective of the evaluation. The following four perspectives were identified:

- single mode (traditional traffic engineering) perspective;
- bimodal transportation system perspective;
- transit system operations perspective; and
- transportation policy/planning perspective.

5.4.1 Single Mode Perspective

The traditional traffic engineering perspective is single modal and is normally operations oriented. The objective is to control and manage traffic in the most effective way within the present environment. From this perspective, mean vehicle travel time is one of the most meaningful performance measures for describing system operation. Vehicle travel time on a specific congested roadway section is a more meaningful measure than vehicle delay at intersections. The reason why it is rarely used in traditional traffic engineering evaluations is twofold:

- Vehicle travel time is correlated with intersection delay and estimating intersection delay is more meaningful because it is based on relevant signal operations parameters. However, as discussed previously, in oversaturated conditions, the traditional intersection delay calculation is inadequate thus it is a very poor estimator of travel time.
- Collecting vehicle travel times is time consuming and could be expensive. However, if the control analysis is carried out by simulation, travel time data collection is not an

issue. In contrast to the traditional delay formula, travel time data can be collected separately for different vehicle categories (e.g. . bus and car).

Vehicle travel time between two points is an aggregate measurement. It includes delays incurred at intersections and the time required to traverse the distance between intersections. For this reason the application of vehicle travel time as a performance measure is tied to particular segments of the transportation system. There are two main aspects that have to be observed when designing travel time data collection for a roadway segment:

- Spatial aspect: The impact of the control must not propagate beyond the travel time measurement points.
- Temporary aspect: Data collection period must be long enough to capture the volume/queue build up and dissipation. The number of vehicles in the system at the beginning and end of the evaluation must be the same for all compared alternatives.

Satisfying the *spatial* aspect seems straightforward. The geographical scope of simulation must cover all the impacts of oversaturation. In practical applications this is often problematic because the growth of queues in oversaturated conditions is very rapid thus confining the scope of analysis may be impossible. In Section 7.3.2.2 an approach to deal with this issue is proposed.

The easiest way to satisfy *the temporary* aspect is to start the simulation with zero vehicles in the system and run the simulation long enough to discharge all vehicles from the system. The downside of this approach is that it does not reflect real-world

conditions. Since the objective is to compare alternative control approaches, deviating from real-world conditions at the beginning and end of the simulation period is not considered to be a major issue.

Travel time collection from microsimulation models is not an issue. Start and end points have to be specified on the network along with the type of vehicles to be logged separately and the time period(s) for data aggregation.

5.4.2 Bimodal Transportation System Perspective

Vehicle travel time, as a performance measure, cannot capture the benefits of multi-modal transportation systems. It was demonstrated in Section 4.3.1.1 that optimization by vehicle-travel time in a bimodal system does not provide the lowest person-travel time.

When bimodal transportation encompasses car and transit travel, mean person travel time is a more meaningful performance measure. Person travel time calculation is based on the vehicle travel time and vehicle occupancy. The determination of car occupancy, is normally not an issue. A default value between 1.1 and 1.3 is reasonable for commuter traffic and field data collection is also a possibility if more accurate data is desired. The variation of transit vehicle occupancy is much greater between urban areas, and within urban areas. Collecting transit occupancy data from the field is problematic, however, many transit agencies systematically collect transit ridership data which can be applied to the calculation. An another option, which has been used in this thesis, is to apply a default occupancy value.

5.4.3 Transit System Operations Perspective

The traditional traffic engineering approach is to use mean delay or mean travel time as the basis of comparison or assessment. This approach is far from perfect even for single mode analysis because it ignores variability. Nevertheless it is accepted as practical approach to measure the impacts of control/congestion on vehicle (car) travel.

Mean delay/travel time is also regularly used to assess the impacts of various control techniques on transit travel times (Dion et al., 2004; Abdulhai et al., 2002). However, using only the mean travel time to assess the impacts of traffic control/management on transit service is inappropriate. Transit vehicles should not be treated as individual/independent vehicles such as personal cars. Transit usually follows a published schedule which is based on historical travel time data. Both the historical mean travel time and its variability are used to determine schedule times. Therefore, transit travel time variability must be included in the evaluation.

It is proposed that the measurement of the 85th percentile transit travel time is used to assess the impacts of various control measures on transit system operation. The 85th percentile value means that 15% of transit vehicles are not expected to complete the trip within the allotted time. This value was selected somewhat arbitrary and other values can be selected based on the particular transit agency's policy. Section 8.9.7 demonstrates the implications of using the 85th percentile transit travel time instead of the mean for evaluation purposes.

5.4.4 Transportation Policy/Planning Perspective

Increasing transit modal share (TMS) by focusing on the development of a multimodal transportation systems is on the top of the agenda of most Canadian municipalities. TMS targets (e.g. 30 % Ottawa, 2003a) are often described as necessary to be able to accommodate future travel demand. TMS is also a good MoP to assess the performance, from a bimodal perspective, of *existing* transportation systems.

However, TMS as a MoP, is inadequate in terms of supporting the design and operation of a transportation system in the transition from existing to future TMS. The inability of the TMS to govern infrastructure design and operation creates a gap between long-term planning/policy objectives and short-term operational objectives.

The evidence of the conflicts between long-term planning and short-term operations objectives are plenty in most urban areas. For example, significant investments are often made in transit signal priority (TSP) supported by long-term plans to increase TMS, but the effectiveness of TSP is regularly limited due to short-term operational considerations. TMS, which is the driving factor for the investment is not used in the performance assessment. Control system design and operation is governed by MoP (e.g. vehicle delay/travel time, V/C) that are totally independent from the TMS. Therefore, operational level decisions are justified based on MoP which may actually decrease TMS. The fault is not with the individual traffic engineering agency, it is with the design of performance measures at the traffic engineering manual and procedures level.

It is proposed that the Transit-Car Travel Time ratio (TCTTr) described by 5.45 is a MoP which can facilitate the connection between the long-term planning objectives, transportation system design criteria, and short-term operational objectives.

$$TCTTr = \frac{IVTT_{bus}}{IVTT_{car}} \quad 5.45$$

$IVTT_{bus}$ – in-vehicle travel time for transit (bus) travel time for the observed section - OD pair (sec)

$IVTT_{car}$ – in-vehicle travel time for car for the observed section - OD pair (sec)

The advantages of TCTTr, as a MoP, are summarized in the following:

- **Deriving the TCTTr is relatively simple:** Travel times for cars and transit vehicles can be collected separately through simulation or field measurements.
- **Connection between TCTTr and TMS is theoretically sound and intuitively acceptable:** Lower TCTTr correlates with higher TMS. TCTTr can be used either as a qualitative measure for the comparison of alternatives, or as a quantitative measure for estimating TMS if the calibration data is adequate. Before TCTTr can be used as a quantitative measure, further research and development is needed.
- **TCTTr is a comprehensive MoP:** It can be used both as a policy objective, system design objective, and operations objective. For example, if the policy objective is to make transit travel times competitive to car travel times within a particular corridor during peak hours (e.g. transit travel time cannot be longer than 1.5 times the vehicle travel times) the transportation system design objective and the control/process

objective are straightforward. TCTTr can be used at all three decision levels. (Figure 5.1).

The new formulation of the TMS calculation does not resolve all the issues. The assumption that the calculated TMS applies to travellers between specific origin-destination pairs is still valid. The technical difficulties of the iterative process of travel time calculation are still there if an analytical approach is pursued. Modal coefficients may need confirmation and/or recalibration for specific locations

5.5 Summary

Vehicle delay, one of the most often used MoP, was designed for the optimization of single mode travel in mostly uncongested conditions. Thus, it has only a limited value in the analysis and control of bimodal systems in congested conditions. Furthermore, the optimization by minimizing vehicle delay leads to the “gaining efficiency by abusing efficiency paradox” which describes a situation where the most efficient element of the system receives the least priority. This leads to the “propagation of inefficiency trap” where the attractiveness of the efficient element (transit travel) diminishes and the long-term decrease of Transit Modal Share (TMS) eventually contributes to the inefficiency of the entire system.

Although the formulation of the Paradox and Trap in this thesis is believed to be new, its repercussions have been already identified by some planning professionals who advocate placing more focus on transit travel vs. car travel at the policy level. However, a major

problem in transportation planning and traffic engineering is that currently there is no connection between the policy level Transit Modal Share (TMS) objective and facility design and operation level objectives. As a result there is disconnect between short-term operations and long-term planning level objectives. To resolve this issue, it is proposed that the formulation of the TMS calculation is reformatted so that it is based on TCTTr (Transit-Car Travel Time ratio).

Measuring the TCTTr is relatively straightforward. It can be carried out for one intersection, a short roadway section or an entire arterial by simulation and/or field measurements. The TCTTr can be directly incorporated into facility planning, design, and operation. For example, if the planning objective is to achieve a 30% TMS in a particular corridor, the required TCTTr can be estimated, and the infrastructure design and operations control/management could be designed accordingly.

The TCTTr is not the only factor that influences TMS but it is one of the few in the domain of infrastructure design and traffic control/management.

6. DEVELOPMENT OF FUZZY LOGIC BASED QUEUE MANAGEMENT

6.1 Introduction

Fuzzy logic techniques have been applied to various transportation related problems. Teodorović et al. (1998) provides a relatively comprehensive summary of transportation problems and fuzzy logic based approaches to solve them. The application of fuzzy logic to queue management is, however, a new approach. The objective of this section is to introduce and explain the concept of fuzzy logic control and to demonstrate its application to a queue management problem.

The section starts with a discussion of the reasons why fuzzy logic is proposed for queue management. This is followed by a discussion of the building blocks of fuzzy logic control. Fuzzy control is then developed for a queue management problem. Finally, tuning techniques are discussed.

6.2 Motivation to Use Fuzzy Logic

There are two main motivations for applying fuzzy logic in queue management. The first motivation is that fuzzy logic can be used for controlling processes which are non linear or not completely understood and described by mathematical models. Vehicle operation and interaction, driver behaviour and its variability are very complex to be modelled by using mathematical models. So far, only a few elements of the transportation system

operation are described by mathematical models (such as queue length, delay, signal coordination etc.) and even those descriptions are valid only for specific conditions. Engineering judgement, in other words heuristic decisions, are routinely utilized in traffic engineering to expand, override, and/or adjust the outputs of the mathematical models. The available mathematical models normally apply to a single mode and they are all vehicle based. Mathematical models that optimize the operation of a bimodal transportation system have not yet been developed.

The second motivation for using fuzzy logic is that it formalizes and integrates engineering judgment directly into the control mechanism. Engineering judgement is usually expressed through symbolic (linguistic) variables and fuzzy logic provides a platform where linguistic variables (from engineering judgement) and numerical variables (from measurements) can be combined into a common control framework.

It is known that, based on relatively simple rules, humans can provide traffic control that can match the performance of sophisticated signal control systems. For example, traffic police can provide effective traffic control at signalized intersections when signals are under-performing due to damage or unforeseen traffic conditions. Similarly, based on relatively simple rules, traffic police could provide effective queue management. The following two general rules provide an overall description of the required control actions:

Rule 1: IF queue length is Short THEN Increase upstream green time

Rule 2: IF queue length is Long THEN Decrease upstream green time

From the rule-base perspective fuzzy logic systems are similar to expert systems. However, there are two significant differences between the fuzzy rule structure and the rule structure in conventional expert systems. First, rules in fuzzy logic represent elastic conditions, while in conventional expert systems rules are rigid. Second, while the classical expert systems rule base is exclusively obtained from human experts, in the case of fuzzy systems, the formation of the rule base can be made by human experts, according to numerical data, or by combining numerical data with the human expert's knowledge (Teodorović, 1998).

6.3 The Elements of Fuzzy Logic Control

The construction and operation of fuzzy control will be explained on a queue management example depicted in Figure 6.1. The objective of queue management is to control vehicle movements by signal operation (M) so that the resulting queue formation achieves predefined performance criteria. The performance criteria can be specified as the queue length at a predefined instant relative to the signal cycle. The developed automatic control combines heuristic elements in the form of fuzzy logic control with the traditional signal control elements.

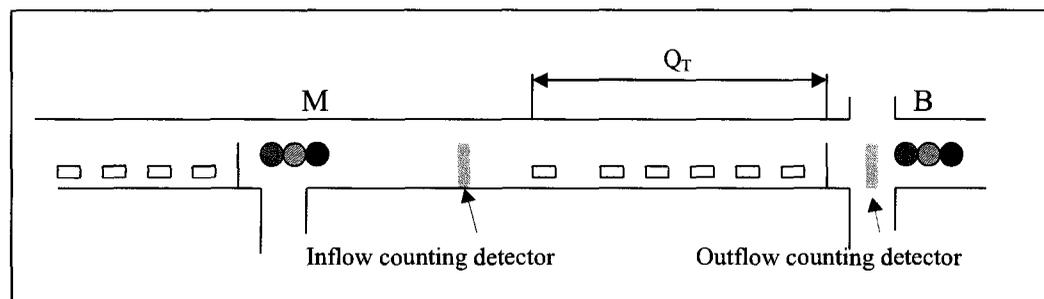


Figure 6.1: Simple Queue Management Set-Up

The block diagram of the queue control system is provided in Figure 6.2. The heuristic component is described by the Fuzzy Block. Fuzzy logic is used to calculate the required green time at the upstream metering intersection. All other signal control elements such as the signal operation at the downstream bottleneck intersection including coordination through an offset, cycle time, the number and sequence of stages, and clearance intervals are determined by using traditional signal control elements (System Operation Block). In addition to signal control elements, the System Operation Block describes vehicle movements and the interactions between them.

The modelling of the traditional elements described by the System Operation Block is modelled by VISSIM, and is not subject to this research. The focus of this section is on the design of the Fuzzy Block.

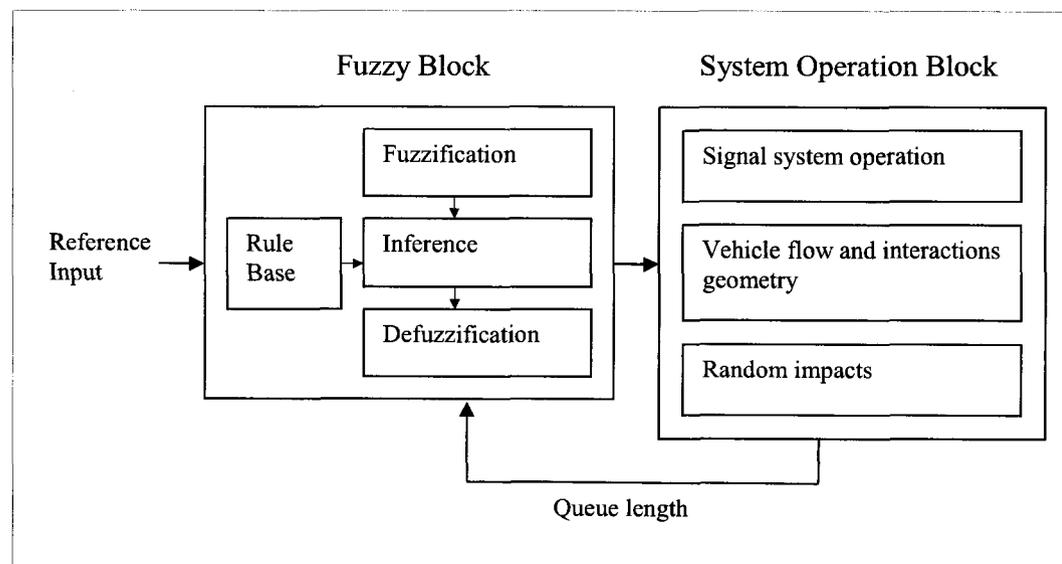


Figure 6.2: Block Diagram of Fuzzy Control Mechanism for Queue Management

The Fuzzy Block represents a non-linear mapping between inputs and outputs. The non-linearity means that the relationship between input values into the Fuzzy Block and the outputs is not necessarily a linear function. The non-linearity is required because the relationship between the System Operation input (green time at the upstream intersection) and the required output (queue length) is not linear, e.g. doubling the green time does not result in 50 % queue length reduction.

The Fuzzy Block by itself represents a static mapping technique (Passino, 1998). This means that if the Fuzzy Block is considered in isolation from the rest of the system the output is determined by the input without any interim feed back. However, in the queue management example (Figure 6.2), the fuzzy inference system is built into a feedback control loop which makes the entire system dynamic. The queue length is measured periodically within the System Block and fed back into the fuzzy inference system, thus providing dynamic control.

The Fuzzy Block contains four elements, the fuzzification of inputs, the rule-base, the inference system, and the defuzzification of outputs. The combination of these four elements is often referred to as the Fuzzy Inference System (FIS). The fundamentals of FIS are explained in the next sections and further details with more background can be found in Yen et al.,(1998), Passino et al., (1997), Teodorović et al., (1998) and Lewis(1997).

6.3.1 Fuzzification

The inputs x_i and outputs y_i into any fuzzy system are real numbers and they are often called *crisp* values. The inputs and outputs belong to crisp sets X_i and Y_i which are often referred to as the *universes of discourse*.

$$x_i \in X_i \quad i = 1 \text{ to } n$$

$$y_i \in Y_i \quad i = 1 \text{ to } m$$

Fuzzy systems mimic the thinking of humans and human logic is based on imprecise definitions. For example, for a driver who is merging from a side street into main street traffic, it is often more meaningful to say that there is an approaching car at a ‘medium’ distance travelling ‘slowly’ than to say that there is a car at 102.8 meters from the intersection travelling at a speed of 5.3 m/sec. Although these linguistic variables or descriptions are not precise quantities, they are sufficient for appropriate driver decisions.

The association of queue lengths (or other fuzzy sets) with specific linguistic variables has two advantages. First, it makes easier to apply human expert’s knowledge in control algorithms. Second, it makes it easier to modify existing control algorithms and develop complex fuzzy logic control systems (Lewis, 1997).

Contrary to mathematical models, fuzzy systems do not rely on the precise measurements of the input variables, they are designed to use fuzzified inputs. Thus, the first step in every FIS is the conversion of the measured or estimated crisp values into fuzzy values.

This process, called *fuzzification*, uses *membership functions* to determine the belonging of crisp values to different linguistic descriptions.

A human expert may describe the queue length with linguistic values such as *Short*, *Medium* and *Long*. A 5 vehicle queue could be defined as a *Short* queue and a 10 vehicle queue as a *Medium* queue. But what about the queue of 8 vehicles? Is it a *Short* or a *Medium* queue? The bivalent logic would require the definition of a precise cut-off point of lets say 7 vehicles. Anything below 7 would be a *Short* queue and anything equal or above 7 vehicles would represent a *Medium* queue. The precise cut-off point however introduces an illogical statement: Six vehicles are describes as a Short queue and the 7 vehicles are described as a Medium queue. Does the difference of only one vehicle really make such a big difference?

The queue description of *Short* or *Medium* determined by the bivalent logic based on a cut-off point represents a very precise but inaccurate definition of the queue length. This approach is normally associated with “machine thinking” which does not recognize the existence of “grey” areas. However, human logic is more comfortable with the approach that a queue of 8 vehicles is somewhere between *Short* and *Medium* but somewhat more *Medium* than *Short*. Mathematically this greyness or fuzziness can be described by degrees to which the queue value belongs to the *Short* and *Medium* linguistic variables at the same time. In this example, the membership in the linguistic variable *Short* is 0.3, and in the linguistic variable *Medium* is 0.7 (Figure 6.3).

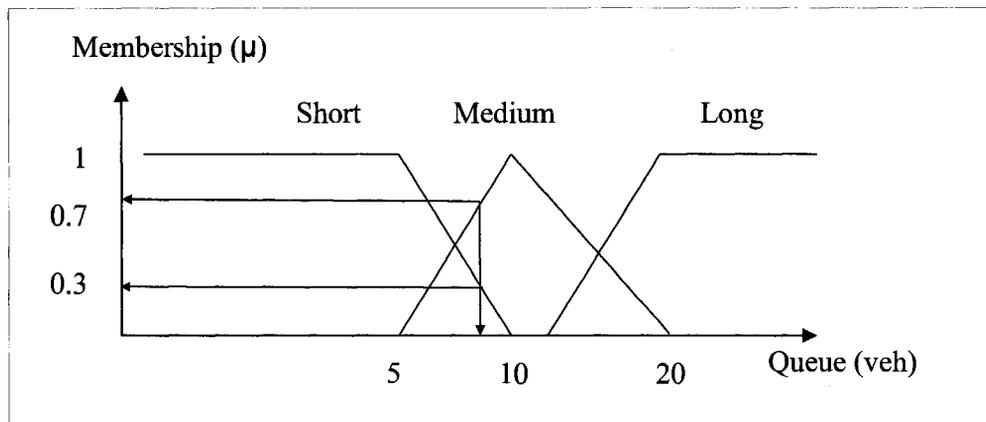


Figure 6.3: Membership Function Examples

The meaning of linguistic terms such as short, medium or long can change when the term is used in different circumstances. For example, on a road sections with shorter signal spacing, the meaning of long queue could be different than on a road section with long signal spacing. Therefore, the definitions of linguistic terms are not universal and they have to be defined for the specific circumstances for which they are applied.

Figure 6.3 depicts three membership functions for the linguistic variables Short, Medium, and Long queues. Although the selection of the form of a membership functions is to a certain degree arbitrary, the membership function has to describe somehow the human perception of the particular quantity. For example, it is quite reasonable to refer to queues of 'x' vehicles or less as short queues. A truncated trapezoid membership function (with a missing left leg) describes this approach. A triangular membership function seems to be reasonable to the human perception for a medium value. As the crisp value moves away from the top of the triangle, the membership in the medium linguistic value decreases and the membership in the short or long linguistic values increases. It is obvious that the

purpose of the membership function is to provide a gradual transition from regions outside of a particular linguistic term (e.g. short queue) to regions inside of a linguistic term.

Membership functions can have different shapes. Beside the aforementioned triangle and truncated trapezoid shapes, they can have a trapezoid, Gaussian bell curve, composite Gaussian bell shape, sigmoid (logistic) curve, polynomial shape, or other shapes drawn by the user. Furthermore, for output values, singleton output membership functions are often used. The singleton membership functions represent a precisely defined value.

Membership functions have to support the control logic and objectives. However, a precise definition of how to construct membership function does not exist. Besides selecting the shape of membership functions, one has to determine its parameters as well.

Fortunately, in most cases, the general shape of the membership function has relatively minor impact on controller performance. Thus, due to their relative simplicity, the triangular and trapezoid membership functions are used the most often. Trapezoid and triangular membership functions can be easily parameterized by 4 and 3 values, respectively. The maximum height is always 1, thus membership values are always between 0 and 1. The advantage of using parameterized membership functions is that they facilitate membership function tuning. Changes to the membership functions, such as widening or narrowing them, can be directly done by modifying the parameters.

Figure 6.4 depicts four membership functions for the linguistic values of Very Small, Small, Zero, and Large. The membership function of the linguistic value Very Small is a trapezoid with a missing left leg (left truncated trapezoid). The four parameters describing the trapezoid are a_{VS} , b_{VS} , c_{VS} , and d_{VS} . The membership function describing the linguistic values from Very Small to Large are defined by Equations 6.1 to 6.4 .

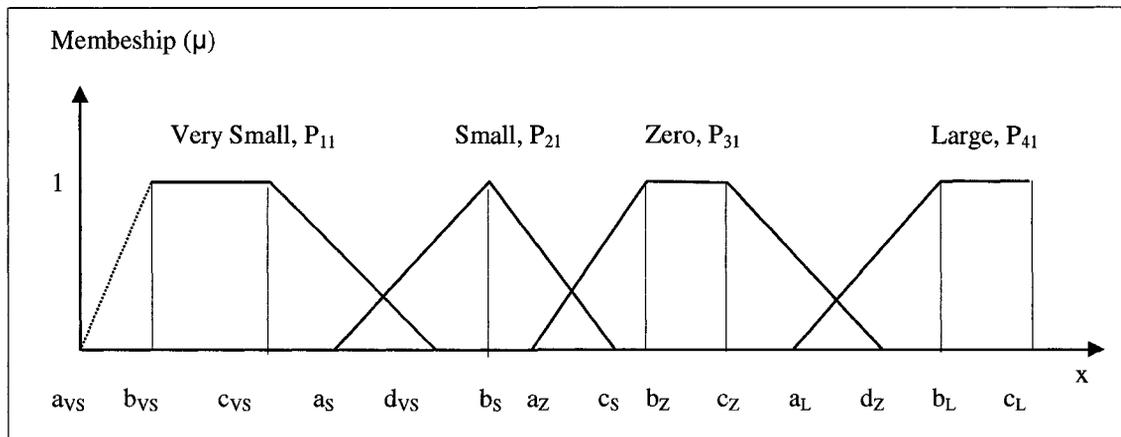


Figure 6.4: Trapezoid and Triangular Membership Functions

The left truncated trapezoid membership function for the linguistic value Very Small (P_{11}) can be described by Equation 6.1.

$$\mu_{VS}(x) = \begin{cases} 1 & \text{if } x \leq c_{VS} \\ \max\left(0, \frac{d_{VS} - x}{d_{VS} - c_{VS}}\right) & \text{if } x > c_{VS} \end{cases} \quad 6.1$$

x – crisp variable

d_{VS}, c_{VS}, \dots - membership function parameters shown on Figure 6.4

The triangular membership function for the linguistic value Small (P_{21}) can be described by Equation 6.2.

$$\mu_S(x) = \begin{cases} \max\left(0, \frac{x-a_S}{b_S-a_S}\right) & \text{if } x \leq b_S \\ \max\left(0, \frac{c_S-x}{c_S-b_S}\right) & \text{if } x > b_S \end{cases} \quad 6.2$$

The trapezoid membership function for the linguistic value Zero (P_{31}) can be expressed by Equation 6.3.

$$\mu_Z(x) = \begin{cases} \max\left(0, \frac{x-a_z}{b_z-a_z}\right) & \text{if } x \leq b_z \\ 1 & \text{if } b_z \leq x < c_z \\ \max\left(0, \frac{d_z-x}{d_z-c_z}\right) & \text{if } x > c_z \end{cases} \quad 6.3$$

The right truncated trapezoid membership function for the linguistic value Large (P_{41}) can be expressed by Equation 6.4.

$$\mu_L(x) = \begin{cases} \max\left(0, \frac{x-a_L}{b_L-a_L}\right) & \text{if } x \leq b_L \\ 1 & \text{if } x > b_L \end{cases} \quad 6.4$$

In this thesis, two approaches to determine the membership function parameters are utilized. In the first approach, the initial membership functions are selected and then they are tuned through a trial-and-error process (Yen et al., 1998). A systematic tuning strategy, often referred to as tuning by scaling is used to expedite the trial-and-error tuning process (Passino et al., 1997). The systematic tuning strategy requires that the membership functions could be parameterized by using a small number of parameters. Further description of the method is provided in Section 6.5.

The second approach is used not only for tuning membership function parameters, but also to determine the rule-base. This approach is based on the process of learning from output data and it is further explored in Chapter 10.

6.3.2 Rule Base

The rule-base is a collection of linguistic rules that describe the control logic. The rules provide a description of what the human expert would do in various circumstances.

Simple control actions can be described by only a few rules, while more complex control actions may require a large number of rules.

Rule bases have two general properties, completeness and consistency (Passino, 1997, pg 61). Completeness means that the rules should cover all possible combinations of input rules, and consistency means that the conclusions of rules should not conflict each other.

Individual rules follow the simple modus ponens or affirmative mode inference which is described as:

IF <antecedent/proposition > THEN <consequent/conclusion>

If antecedent is true, the consequent will also be true. In mathematical logic, the modus ponens inference mechanism is not the only reasoning technique. However, most human reasoning follows the affirmative mode concept, and fuzzy systems have been based on that concept almost exclusively.

Usually the inputs of the fuzzy system represent the antecedents and the outputs represent the consequents. The IF –THEN rules can be represented in many forms. The simplest rule has single input and single output (SISO) form. For example, the rule base may contain the following rule:

Rule 1: IF queue is Short, THEN increase green time at the upstream intersection

The above rule has only one antecedent which is the queue length and only one consequent which is the green time change. This is called an atomic proposition (Pant et al., 2007).

In fuzzy systems, similarly to human logic, decisions are often based on more than one input. The following rule base includes three rules and each rule has a compound proposition consisting of two input values thus the name MISO (multi input-single output):

Rule 1: IF queue is Short AND queue is Increasing THEN reduce green time a Little
OR

Rule 2: IF queue is Just Right AND queue is not Changing THEN no Change
OR

Rule 3: IF queue is Long AND queue is Increasing THEN decrease green time a Lot

The third form, abbreviated as MIMO (multi input - multi output), is similar to MISO but it has more than one output (Passino, 1998).

The propositions in individual rules (MISO or MIMO form) and the rules are connected with logical operators. The AND operator is used to connect individual propositions

within the rule and the OR operator is used to connect different rules. The AND and OR operators are not the only operators that are used in fuzzy systems, there are many others. The selection of logical operators has to describe the human expert's logic and support the control objective.

Fuzzy reasoning can be viewed as a superset of the traditional Boolean logic (MathWorks, 2001). If the fuzzy membership values are at their extremes of 0 and 1, fuzzy logic corresponds to the Boolean logic. Assume that A and B represent two propositions in both Boolean (binary) and fuzzy logic. In the binary logic, their values could be 0 or 1. If we assume that in the fuzzy logic, propositions A and B have only extreme membership values (0 or 1), the traditional truth table is identical for the Boolean and fuzzy logic Table 6.1.

Table 6.1: Truth Values Table for Fuzzy Logic Membership Values at their Extreme and Truth Values for Boolean Logic

AND Operator			OR Operator			NOT Operator	
A	B	A AND B	A	B	A OR B	A	NOT A
0	0	0	0	0	0	0	1
0	1	0	0	1	1	1	0
1	0	0	1	0	1		
1	1	1	1	1	1		

In Boolean (or bivalent) logic, propositions A and B can have only two values 1 (TRUE) or 0 (FALSE), therefore Table 6.1 provides a complete description of the logic. This is not the case in fuzzy logic. In fuzzy logic, propositions A and B are graded by their membership functions and can have values between 0 and 1, thus a different interpretation of the logical operators AND, OR, and NOT are required. An interpretation

of fuzzy logic operators has been provided by Zadeh (1965). The fuzzy logic equivalent of the AND is the MIN (minimum) operator, the equivalent of the OR is the MAX (maximum) operator, and the equivalent of NOT A is the 1-A (additive complement) operator.

For the extreme cases, the fuzzy operators provide exactly the same outcomes as the Boolean operators Table 6.1. However, for propositions that have membership function between 0 and 1, the fuzzy operator can be also interpreted. For example, if the truth value of A is 0.3 and the truth value of B is 0.7, the outcome of the Fuzzy AND (MIN) operator is 0.3, and the outcome of the fuzzy OR (MAX) operator is 0.7 (last row in Table 6.2).

Table 6.2: Fuzzy Operators AND, OR and NOT

Fuzzy AND (MIN) Operator			Fuzzy OR (MAX) Operator			Fuzzy NOT (1-A) Operator	
A	B	MIN(A,B)	A	B	MAX(A,B)	A	1-A
0	0	0	0	0	0	0	1
0	1	0	0	1	1	1	0
1	0	0	1	0	1		
1	1	1	1	1	1		
0.3	0.7	0.3	0.3	0.7	0.7	0.3	0.7

The fuzzy AND operator is representing the intersection of two fuzzy sets and the fuzzy OR operator is representing the union of two fuzzy sets. This can be described by the following equations:

The Fuzzy AND Operator

$$\mu_{A \cap B}(x) = \text{MIN}\{\mu_A(x), \mu_B(x)\}$$

6.5

x - crisp value

A, B - linguistic variables, $A = \{\text{very short, short, zero, long, ...}\}$

$\mu_A(x)$ - membership of x in linguistic variable A

The Fuzzy OR Operator

$$\mu_{A \cup B}(x) = \text{MAX}\{\mu_A(x), \mu_B(x)\} \quad 6.6$$

The Fuzzy NOT Operator

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x) \quad 6.7$$

The AND, OR, and NOT fuzzy operators are the most often used logical operators but they are not the only operators. Algebraic product, boundary product, Hamacher's MIN operator, Yager's MIN operator, Dubois and Prade's MIN operator are some of the other logical operators describing the intersections of fuzzy sets (Teodorović, 1998). Similarly, the algebraic sum, bounded sum, Hamacher's MAX operator, Yager's MAX operator, Dubois and Prade's MAX operator are some of the other logical operators describing the union of fuzzy sets (Teodorović, 1998).

6.3.3 Inference and Defuzzification

The inference mechanism or (fuzzy reasoning) can be divided into two parts. The first part includes the evaluation of the importance or firing strengths of each individual rule. The second part is the combination of the outcomes of each rule into a single output. The output can be either a new fuzzy set or a crisp value. If the output is a fuzzy set, the final crisp output value is generated by defuzzification.

Assume that the fuzzy system has two inputs and one output and the rule base consists of the following rules:

Rule 1: If x_1 is P_{11} and x_2 is P_{12} THEN y_1 is R_1

OR

Rule 2: If x_1 is P_{21} and x_2 is P_{22} THEN y_2 is R_2

OR

Rule k: If x_1 is P_{k1} and x_2 is P_{k2} THEN y_3 is R_k

x_i - crisp inputs of variable X_i

P_{ij} - Membership functions j for variable x_i

R_i - Output membership function i

y_i - output of rule i

First, the importance or firing strength or weight of each rule is evaluated. This describes the membership of the output y_i in the membership function R_i . If the AND operator is used to connect various inputs, the MIN operator is used to find the lowest membership value as described in Equations 6.8 below. The justification of using the MIN operator is the following. The membership functions for each antecedent represent the degree of truth of that claim, and the truth of the entire rule cannot be higher than the truth value of the lowest contributing element of the rule.

The strength or importance of individual rules is calculated by Equation 6.8.

$$\mu_{R_1} = w_1 = \text{MIN}[\mu_{P_{11}}(x_1), \mu_{P_{12}}(x_2), \dots, \mu_{P_{1n}}(x_n)] \quad 6.8$$

$$\mu_{R2} = w_2 = \text{MIN}[\mu_{P21}(x_1), \mu_{P22}(x_2), \dots, \mu_{P2n}(x_n)]$$

...

$$\mu_{Rk} = w_k = \text{MIN}[\mu_{Pk1}(x_1), \mu_{Pk2}(x_2), \dots, \mu_{Pkn}(x_n)]$$

μ_{Rk}, w_k - output membership or rule k often referred to as the weight or firing strength

$\mu_{P11}(x_1)$ - membership of crisp variable x_1 in linguistic category P_{11}

The consequent R of the rules can have the following three different general forms (Yen et al., 1998):

- R is a fuzzy set
- R has a functional form described by 6.9, and
- R is a crisp (non-fuzzy, constant) value.

$$y = a_0 + \sum_{i=1}^n (a_i \times x_i) \tag{6.9}$$

y – output variable of the rule

x_i – crisp input variables

a_0, a_i – constants

The next part in the fuzzy inference process is the combination of outputs of all rules and the conversion of the fuzzy outputs into crisp outputs. The inference process is somewhat different for the three forms identified above.

Figure 6.5 shows the combination of rules when the output is defined by a fuzzy set. This process is often referred to as the Mamdani process (Mamdani et al., 1975). R_1 and R_2 represent output membership functions and μ_1 and μ_2 represent the firing strength of individual rules.

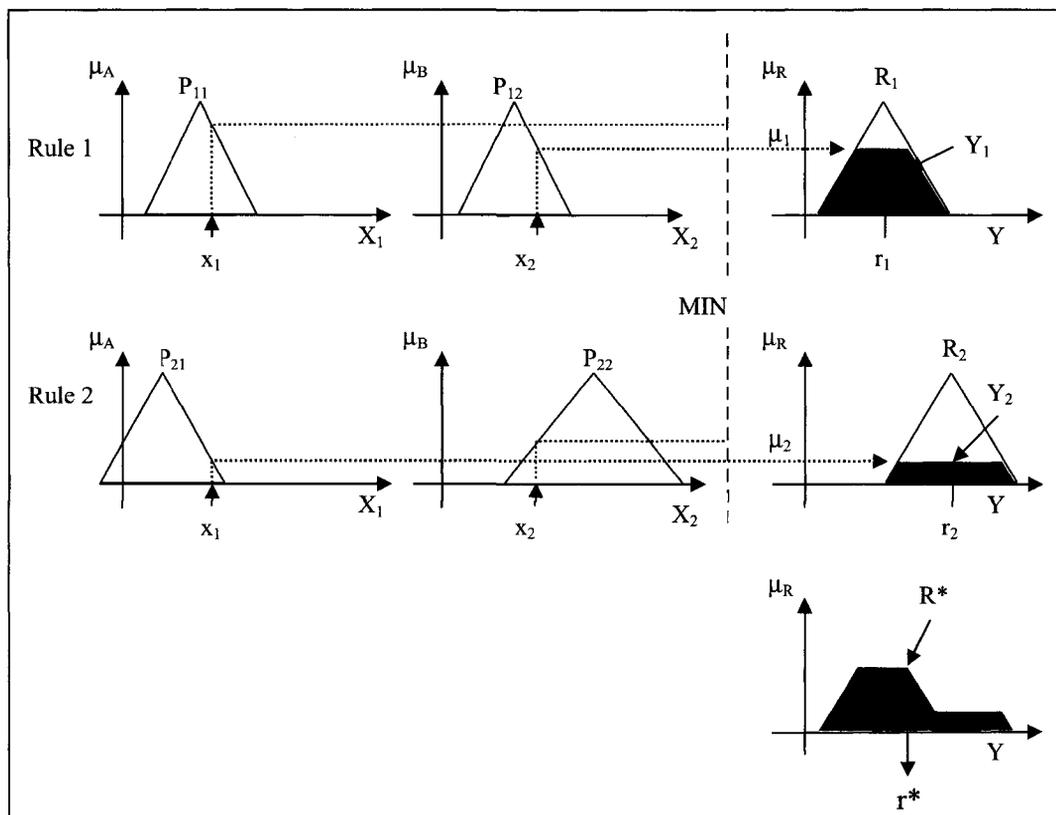


Figure 6.5: Graphical Presentation of the Inference Mechanism when the Output is a Fuzzy Set

The output of the inference process is defined by the contour of a membership function R^* (Figure 6.5). The objective of defuzzification is to produce a single crisp output that can be used in process control. The crisp output should be a value that best represents the contour. There are many different approaches to defuzzification (MathWorks, 2001).

A frequently used approach is the “center of gravity” COG method described by Equation 6.10.

$$r^* = \frac{\sum_i r_i \int \mu_i(Y_i)}{\sum_i \int \mu_i(Y_i)} \quad 6.10$$

r^* - crisp output value

r_i – center of the membership function i

μ_i – fuzzy output function for rule i

Figure 6.6 shows the combination of rules when the output is defined by the functional form expressed in Equation 6.9. This type of inference is often called Takagi-Sugeno type inference since it was developed by these two researchers (Sugeno et al., 1985). The output of individual rules is calculated from two pieces of information, the center of the output and the firing strength of the rule. The centre of the output can be represented by the linear combination of the crisp input values x_i expressed by Equation 6.9. The strength of the rule can be expressed by Equation 6.8. To combine the outputs of all rules, the weighted average of individual rules is calculated by using Equation 6.11.

$$r^* = \frac{\sum_i y_i \mu_i}{\sum_i \mu_i} \quad 6.11$$

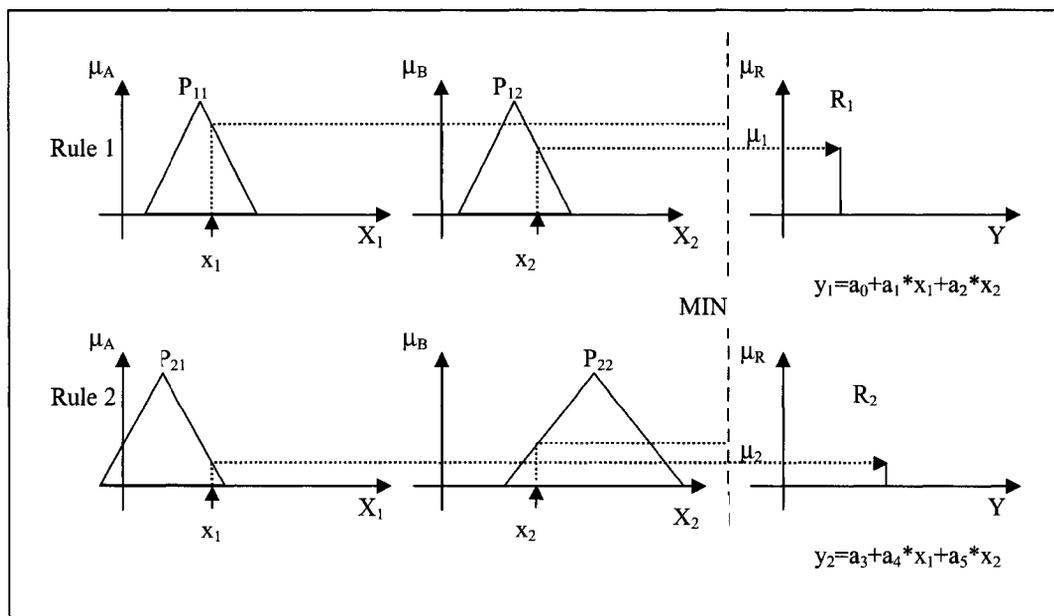


Figure 6.6: Graphical Presentation of the Inference Mechanism when the Output is a Function

Figure 6.7 shows the third method for calculating the consequent which is a special case of the Takagi and Sugeno type inference. In this case the output functions are represented by singleton values where the output centres are constants a_1 and a_2 . Similarly to the previous two cases, the rule strengths are determined by Equation 6.8. The final output is determined by calculating the weighted average (Centre Average defuzzification method) of the outputs of individual rules by using 6.12.

$$r^* = \frac{\sum_i a_i \mu_i}{\sum_i \mu_i} \quad 6.12$$

a_i – centre of singleton output membership function (constant)

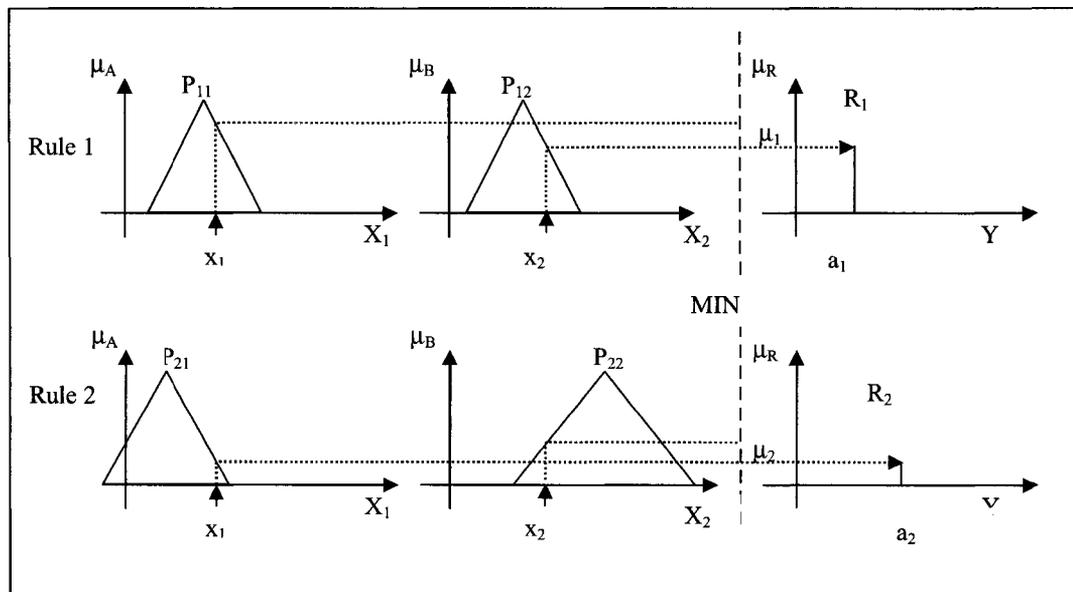


Figure 6.7: Graphical Presentation of the Inference Mechanism when the Output Function is a Singleton Value

Because the shape of the output membership function does not play any role, it makes sense to use the Centre Average defuzzification method when the output membership functions are defined as singleton values.

Expert systems are also based on similar IF-THEN rule structures but they are conceptually and structurally different from fuzzy systems. Most conventional expert systems have binary rules and they evaluate rules by following a predefined hierarchical structure, every rule has a predefined number of outcomes and rules are executed by following the route tree. Fuzzy systems evaluate all rules in the database at the same time and combine the conclusions of all rules in the rule base through an inference mechanism.

6.4 Development of Queue Management Based on Fuzzy Logic

The demonstration of the steps required to build a fuzzy control system are carried out on a two-input one output fuzzy controller that provides queue management functionality.

The control objective is to maintain a target queue length of Q_T at time t at the approach to intersection B where t represents relative time within the signal cycle (Figure 6.1). In the general case, the target queue length is not constant. It can change from cycle k to cycle $k+1$ as higher order control objectives are executed. For example, if a bus is detected upstream, the Q_T could be shorter.

The design of the control system starts with the expert's definition of variables that he needs to watch to be able to control the formation of queues. The fuzzification process converts crisp input variables into linguistic terms that can be utilized by the rule-base. The rule-base describes the human expert's knowledge of the desired control in linguistic terms. The inference mechanism is another description of the human expert's knowledge; it specifies the importance of each individual rule and the way how multiple rules are combined together. The outcome of the inference mechanism is a fuzzy output, which is converted by the defuzzification mechanism to a crisp input that can be used by the process.

6.4.1 Definition of Input and Output Variables

The first variable that is needed to manage queue length is the difference between the target queue length Q_T and the measured queue length Q_M . The variable describing the difference is called the *error*, $e(k)$ where the index k represents the signal cycle when the queue sample was taken. It is obvious that to be able to manage queue length, it is important to know how far is the system from the target value.

The second variable describes how fast and in what direction does the error change. On signalised intersection approaches, the queue length is constantly changing during the signal cycle. This change is, however, not very relevant from the queue management perspective. For control purposes we are not measuring the queue length continuously, but in time intervals which correspond to the cycle length of the bottleneck intersection. Therefore, the *change in error*, Δe describes the difference in the error value between the two measurements in two consecutive signal cycles, k and $k+1$.

$$e(k) = Q_T - Q_M(k) \quad 6.13$$

$$\frac{\Delta e(k)}{\Delta k} = \frac{e(k) - e(k-1)}{1} \quad 6.14$$

The next step is to describe the output of the control system. Assume that the system controlling variable is the duration of green time at the upstream metering intersection M (Figure 6.1).

There could be a number of choices in selecting the controlling variable. It could be the green/cycle ratio at the metering intersection or it could be the duration of green time or the green/cycle ratio for the bottleneck intersection. The green duration at the metering intersection was chosen because objectives other than the target queue length have to be met as well. These objectives include the requirements to have a common cycle length for the metering and bottleneck intersection and the requirement of not to interfere with the operation of the bottleneck intersection. Adjacent intersections are often operating at a common cycle length because it simplifies signal coordination. Since the queue management approach developed in this thesis is an add-on to existing signal timing procedures, it is required that the queue management algorithms are implemented only at key individual locations without the need to re-time signals at adjacent intersections.

Example: Suppose that the target queue length measured at the beginning of the green phase in cycle k is $Q_T(k)=10$ vehicles and the measured queue length is $Q_M(k)=17$ vehicles. The queue length in the previous cycle ($k-1$) was $Q_M(k-1)=19$ vehicles. By using Equation 6.13 the error is $e(k)=10-17=-7$ and by using Equation 6.14 the change of error is $\Delta e(k)=17-19=-2$.

6.4.2 Fuzzification

The rules describing the operation of the control system are based on the expert human's logic and they are expressed in linguistic terms. For example, the expert would look at the queue and it would likely register it as shorter than expected, longer than expected, or approximately as long as expected. Furthermore, the decision in term of the duration of

the green time does not require that the expert knows exactly how many vehicles are in the queue, his thinking is based on the information that the queue is longer or shorter than what it is expected.

The linguistic value of 'longer than expected' is described by the expression *negative error*. Negative error is further specified as *Negative Large* or *Negative Small* error.

Similarly, the positive error is specified as *Positive Small* and *Positive Large*. Table 6.3 and Table 6.4 provide the membership function parameters for the *error, e* and *change of error, Δe* input variables. Figure 6.8 provides a graphical presentation of the membership functions of the input variable *error*.

Table 6.3: Parameter Values of Membership Functions of the Error, e

Linguistic Variable	Error, e			
	a	b	c	d
Negative Large, P ₁₁	-99	-95	-10	-5
Negative Small, P ₂₁	-10	-5	0	
Zero, P ₃₁	-5	0	5	
Positive Small, P ₄₁	0	5	10	
Positive Large, P ₅₁	5	10	95	99

Table 6.4: Parameter Values of Membership Functions of the Change of Error, Δe

Linguistic Variable	Change of Error, Δe			
	a	b	c	d
Negative Large, P ₁₂	-99	-95	-5	-2.5
Negative Small, P ₂₂	-5	-2.5	0	
Zero, P ₃₂	-2.5	0	2.5	
Positive Small, P ₄₂	0	2.5	5	
Positive Large, P ₅₂	2.5	5	95	99

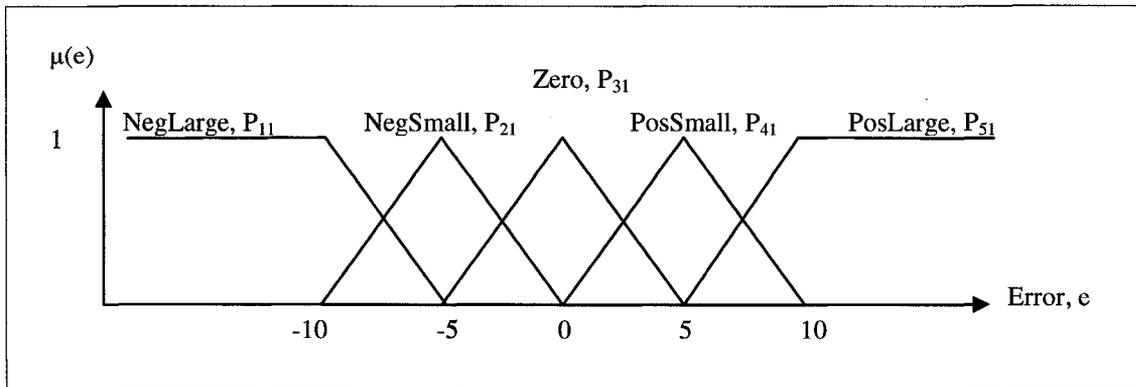


Figure 6.8: Membership Functions

The membership functions of the output variable are similarly defined. When defining the output membership function it is important to know what defuzzification method will be used. For example, if the Center of Gravity (COG) defuzzification method is used, membership functions must be convex (otherwise the area under the membership function is infinite) (Passino, 1998). Truncated trapezoid, sigmoidal or other non-convex membership functions cannot be used.

In this example, the output variable is the change of the green time duration, and five linguistic values are used, Negative Large, Negative Small, Zero, Positive Small, and Positive Large. All five membership functions have a triangular shape. The parameters of the membership function for the output linguistic variable are provided in Table 6.5.

Table 6.5: Parameter Values of Membership Functions for the Output Linguistic Value of Green Time Change

Linguistic Variable	Green Time Change, R		
	a	b	c
Negative Large, P ₁₁	-12	-8	-4
Negative Small, P ₂₁	-8	-4	0
Zero, P ₃₁	-4	0	4
Positive Small, P ₄₁	0	4	8
Positive Large, P ₅₁	4	8	12

Example:

Calculating memberships of the error input variable, $e(k) = -7$:

Membership in the Negative Large category is calculated by using equation (6.1) and data from Table 6.3:

$$\mu_{NL}(e) = \frac{-5 - (-7)}{-5 - (-10)} = \frac{2}{5} = 0.4$$

Membership in the Negative Small category is calculated by using equation (6.2) and data from Table 6.3:

$$\mu_{NS}(e) = \frac{-7 - (-10)}{-5 - (-10)} = \frac{3}{5} = 0.6$$

Membership in the Zero, Positive Small, and Positive Large categories is 0 according to calculations based on 6.3 and 6.1.

Calculating memberships of the change of error input variable, $\Delta e(k) = 2$:

Memberships are similarly calculated by using the same equations and membership function parameters provided in Table 6.4. Results for both input variables are provided in Table 6.6:

Table 6.6: Membership Function Values

Linguistic category	$\mu(e)$	$\mu(\Delta e)$
Negative Large	0.4	0.0
Negative Small	0.6	0.0
Zero	0	0.2
Positive Small	0.0	0.8
Positive Large	0.0	0.0

To simplify programming and to be able to generate a generic fuzzy logic algorithm code, the input variables are normalized and normalized membership functions are used. The normalization of input variable is carried out by multiplying the input variables with the normalization factor defined by 6.15.

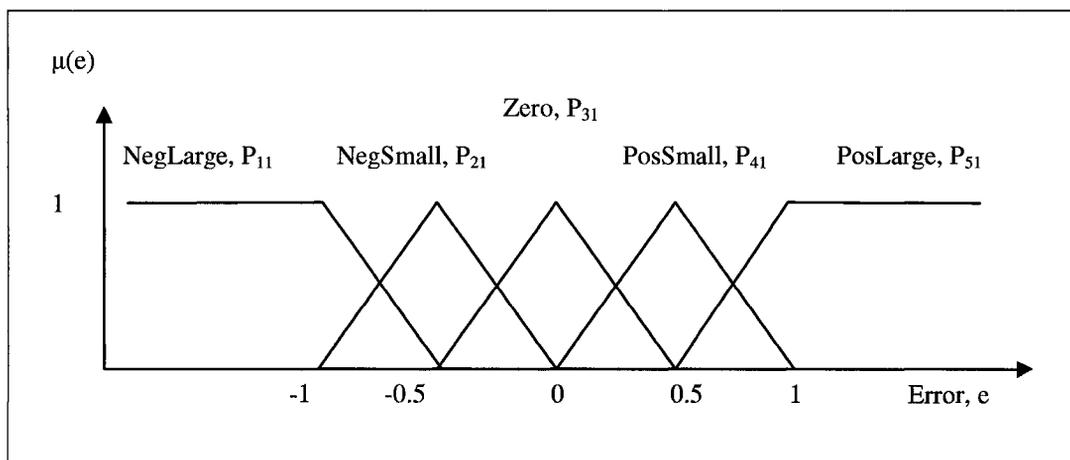
$$n_i = \frac{1}{\text{MAX}(\text{ABS}(\text{input variable}))} \quad 6.15$$

n_i -normalization factor for input variable i

Membership functions are normalized by changing the scale of the x-axis by multiplying all values in the universe of discourse by the normalization factor n_i . The parameters of the normalized membership functions for the input variable *error* are presented in Table 6.7 and they are depicted in Figure 6.9.

Table 6.7: Normalized Parameter Values of Membership Functions for Error, e

Linguistic Variable	Membership function parameters for $error, e$			
	a	b	c	d
Negative Large, P_{11}	-99	-95	-1.0	-0.5
Negative Small, P_{21}	-1.0	-0.5	0	
Zero, P_{31}	-0.5	0	0.5	
Positive Small, P_{41}	0	0.5	1.0	
Positive Large, P_{51}	0.5	1.0	95	99

**Figure 6.9: Normalized Membership Functions of Error, e**

6.4.3 Definition of the Rule-Base

The rule-base is a collection of linguistic rules that describe the control logic. The rules provide a description of what the human expert would do in various circumstances. For example the control logic can be described as follows: If the measured queue length is shorter than the target queue length ($e > 0$) and the queue length is getting shorter compared to the previous cycle ($\Delta e > 0$) then the control action should be to provide more green time at the metering intersection.

This control logic is easy to understand but not sufficiently defined; e , Δe , can have different values thus the control action 'increase green time' should have different levels as well. Therefore, the rule base includes all possible combinations of the membership categories of the *error* and *change of error* variables. Both input variables have five membership categories and the number of rules that covers all possibilities is 25 (i.e. 5^2).

To simplify the definition of rules, abbreviations of linguistic numeric values are often used. For example, the linguistic value Negative Large can be described as NL or by the linguistic-numeric variable '-2' as described in Table 6.8. The linguistic-numeric values do not represent a quantity they only show the sign of the input/output value and indicate the size in relation to the other linguistic values. In the rest of the thesis all three rule descriptions will be used as appropriate.

Table 6.8: Various Descriptions of Linguistic Values

Linguistic variable	Abbreviation	Linguistic-numeric value
Negative Large	NL	-2
Negative Small	NS	-1
Zero	Z	0
Positive Small	PS	1
Positive Large	PL	2

A complete list of rules for the two input variables, e and Δe , is presented in Table 6.9.

Table 6.9: List of Rules for Two Input and One Output Variable

Description of Rules
Rule 1: IF <i>error</i> is NL and <i>change in error</i> is NL THEN <i>green time change</i> is NL
Rule 2: IF <i>error</i> is NL and <i>change in error</i> is NS THEN <i>green time change</i> is NL
Rule 3: IF <i>error</i> is NL and <i>change in error</i> is Z THEN <i>green time change</i> is NS
Rule 4: IF <i>error</i> is NL and <i>change in error</i> is PS THEN <i>green time change</i> is NS
Rule 5: IF <i>error</i> is NL and <i>change in error</i> is PL THEN <i>green time change</i> is Z
Rule 6: IF <i>error</i> is NS and <i>change in error</i> is NL THEN <i>green time change</i> is NL
Rule 7: IF <i>error</i> is NS and <i>change in error</i> is NS THEN <i>green time change</i> is NS
Rule 8: IF <i>error</i> is NS and <i>change in error</i> is Z THEN <i>green time change</i> is NS
Rule 9: IF <i>error</i> is NS and <i>change in error</i> is PS THEN <i>green time change</i> is Z
Rule 10: IF <i>error</i> is NS and <i>change in error</i> is PL THEN <i>green time change</i> is PS
Rule 11: IF <i>error</i> is Z and <i>change in error</i> is NL THEN <i>green time change</i> is NS
Rule 12: IF <i>error</i> is Z and <i>change in error</i> is NS THEN <i>green time change</i> is Z
Rule 13: IF <i>error</i> is Z and <i>change in error</i> is Z THEN <i>green time change</i> is Z
Rule 14: IF <i>error</i> is Z and <i>change in error</i> is PS THEN <i>green time change</i> is Z
Rule 15: IF <i>error</i> is Z and <i>change in error</i> is PL THEN <i>green time change</i> is PS
Rule 16: IF <i>error</i> is PS and <i>change in error</i> is NL THEN <i>green time change</i> is NS
Rule 17: IF <i>error</i> is PS and <i>change in error</i> is NS THEN <i>green time change</i> is Z
Rule 18: IF <i>error</i> is PS and <i>change in error</i> is Z THEN <i>green time change</i> PS
Rule 19: IF <i>error</i> is PS and <i>change in error</i> is PS THEN <i>green time change</i> PS
Rule 20: IF <i>error</i> is PS and <i>change in error</i> is PL THEN <i>green time change</i> PL
Rule 21: IF <i>error</i> is PL and <i>change in error</i> is NL THEN <i>green time change</i> is Z
Rule 22: IF <i>error</i> is PL and <i>change in error</i> is NS THEN <i>green time change</i> is NS
Rule 23: IF <i>error</i> is PL and <i>change in error</i> is Z THEN <i>green time change</i> PL
Rule 24: IF <i>error</i> is PL and <i>change in error</i> is PS THEN <i>green time change</i> PL
Rule 25: IF <i>error</i> is PL and <i>change in error</i> is PL THEN <i>green time change</i> PL

The same rule base in a different, more convenient format, is provided in Table 6.10.

Table 6.10: Control Rule Table

Green time adjustment			Change of effort, Δe				
			Queue increasing		Queue stable	Queue decreasing	
			-2	-1	0	1	2
Error, e	Longer than target	-2	-2	-2	-1	-1	0
		-1	-2	-1	-1	0	1
	At target	0	-1	0	0	0	1
	Shorter than target	1	-1	0	1	1	2
2		0	1	2	2	2	

The rule-base is not unique, it can have many variations. However, for the control system to achieve its objectives, the rules describing the control system must be fundamentally correct.

The rule-base is determined either by capturing the knowledge of experts who understand the operation of the system or by using automated learning techniques. The latter approach requires sufficient amount of input-output data points and more details are provided in Chapter 10.

6.4.4 Inference and Defuzzification

As described earlier, the inference mechanism is a mathematical representation of the human expert's knowledge on how to control the process. This mechanism can be divided into two distinctive steps. In the first step the importance of every individual rule is quantified. In the second step, the recommendations of various rules are combined.

In the first step the MIN fuzzy operator is used (Equation 6.8). The overall justification for using the MIN (AND) compositional rule of inference is the fact that we cannot be more certain about our conclusions than we are certain about individual premises. For example, if the first premise *<error is Negative Large>* has a truth value of $\mu(e)=0.4$, and the second premise *<change of error is Negative Large>* has a truth value of $\mu(\Delta e)=0$, the MIN (AND) composition of the two premises cannot be higher than the value of the lower truth value.

By using Equation 6.8, we can calculate the weight or importance of each rule. The weight of Rule 1 is:

$$w_1 = \min [\mu_{NL}(e=-7), \mu_{NL}(\Delta e=2)] = \min [0.4, 0] = 0$$

The consequent of Rule 1 is <green time change is Negative Large> and the weight of this rule is 0.4, thus the truth value of the instruction to provide a Negative Large change to the green time is 0 (i.e. $\mu_{NL}(Y_1)=0$).

The inference and defuzzification is demonstrated on the following numerical example:

The truth values of the antecedents and the firing strengths of the consequents are provided for all rules in Table 6.11. Only Rules 3, 4, 8, and 9 have firing strengths that are different from 0.

The second step of the inference process is the combination of all 25 rules. In fuzzy logic, all rules are evaluated every time, and the sequence of rules is not important. Rules are always connected with the OR operator. The result of the union is a new fuzzy set and it represents the union of all contributing rules.

The following graphical example shows the inference process for rules 3, 4, 8, and 9. The other rules have strength of zero, and they do not influence the outcome.

Table 6.11: Rule Base for the Queue Management Example

Rule #	First premise	Operator	Second premise	Consequent
1	$\mu_{NL}(-7)=0.4$	AND	$\mu_{NL}(2)=0$	$\mu_{NL}(Y_1)=0$
2	$\mu_{NL}(-7)=0.4$	AND	$\mu_{NS}(2)=0$	$\mu_{NL}(Y_2)=0$
3	$\mu_{NL}(-7)=0.4$	AND	$\mu_Z(2)=0.2$	$\mu_{NS}(Y_3)=0.2$
4	$\mu_{NL}(-7)=0.4$	AND	$\mu_{PS}(2)=0.8$	$\mu_{NS}(Y_4)=0.4$
5	$\mu_{NL}(-7)=0.4$	AND	$\mu_{PL}(2)=0$	$\mu_Z(Y_5)=0$
6	$\mu_{NS}(-7)=0.6$	AND	$\mu_{NL}(2)=0$	$\mu_{NL}(Y_6)=0$
7	$\mu_{NS}(-7)=0.6$	AND	$\mu_{NS}(2)=0$	$\mu_{NS}(Y_7)=0$
8	$\mu_{NS}(-7)=0.6$	AND	$\mu_Z(2)=0.2$	$\mu_{NS}(Y_8)=0.2$
9	$\mu_{NS}(-7)=0.6$	AND	$\mu_{PS}(2)=0.8$	$\mu_Z(Y_9)=0.6$
10	$\mu_{NS}(-7)=0.6$	AND	$\mu_{PL}(2)=0$	$\mu_{PS}(Y_{10})=0$
11	$\mu_Z(-7)=0$	AND	$\mu_{NL}(2)=0$	$\mu_{NS}(Y_{11})=0$
12	$\mu_Z(-7)=0$	AND	$\mu_{NS}(2)=0$	$\mu_Z(Y_{12})=0$
13	$\mu_Z(-7)=0$	AND	$\mu_Z(2)=0.2$	$\mu_Z(Y_{13})=0$
14	$\mu_Z(-7)=0$	AND	$\mu_{PS}(2)=0.8$	$\mu_Z(Y_{14})=0$
15	$\mu_Z(-7)=0$	AND	$\mu_{PL}(2)=0$	$\mu_{PS}(Y_{15})=0$
16	$\mu_{PS}(-7)=0$	AND	$\mu_{NL}(2)=0$	$\mu_{NS}(Y_{16})=0$
17	$\mu_{PS}(-7)=0$	AND	$\mu_{NS}(2)=0$	$\mu_Z(Y_{17})=0$
18	$\mu_{PS}(-7)=0$	AND	$\mu_Z(2)=0.2$	$\mu_{PS}(Y_{18})=0$
19	$\mu_{PS}(-7)=0$	AND	$\mu_{PS}(2)=0.8$	$\mu_{PS}(Y_{19})=0$
20	$\mu_{PS}(-7)=0$	AND	$\mu_{PL}(2)=0$	$\mu_{PL}(Y_{20})=0$
21	$\mu_{PL}(-7)=0$	AND	$\mu_{NL}(2)=0$	$\mu_Z(Y_{21})=0$
22	$\mu_{PL}(-7)=0$	AND	$\mu_{NS}(2)=0$	$\mu_{NS}(Y_{22})=0$
23	$\mu_{PL}(-7)=0$	AND	$\mu_Z(2)=0.2$	$\mu_{PL}(Y_{23})=0$
24	$\mu_{PL}(-7)=0$	AND	$\mu_{PS}(2)=0.8$	$\mu_{PL}(Y_{24})=0$
25	$\mu_{PL}(-7)=0$	AND	$\mu_{PL}(2)=0$	$\mu_{PL}(Y_{25})=0$

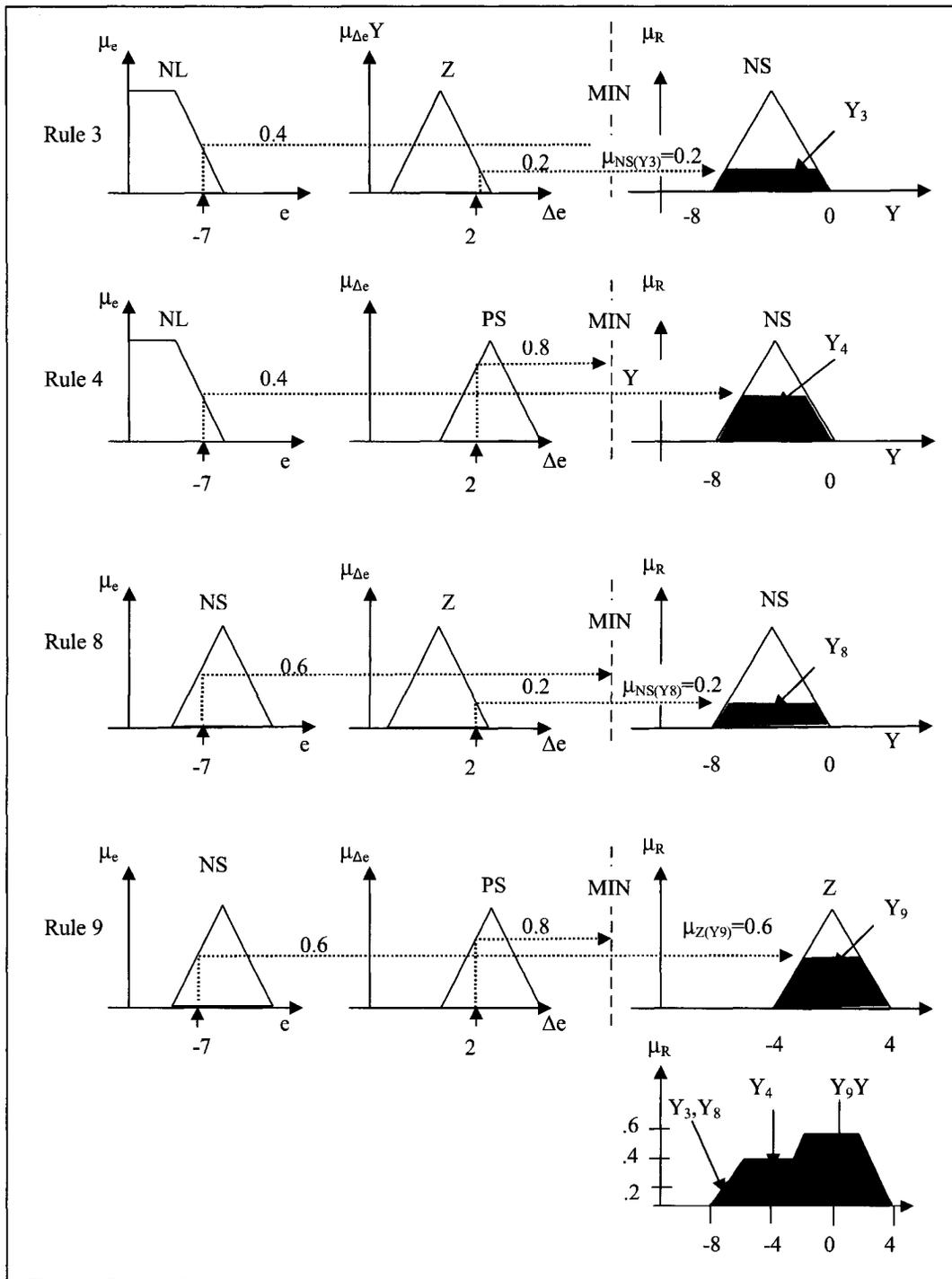


Figure 6.10: Graphical Presentation of the Mamdani Inference Process

The result of the rule combination is shown at the bottom of Figure 6.10 and repeated in Figure 6.11.

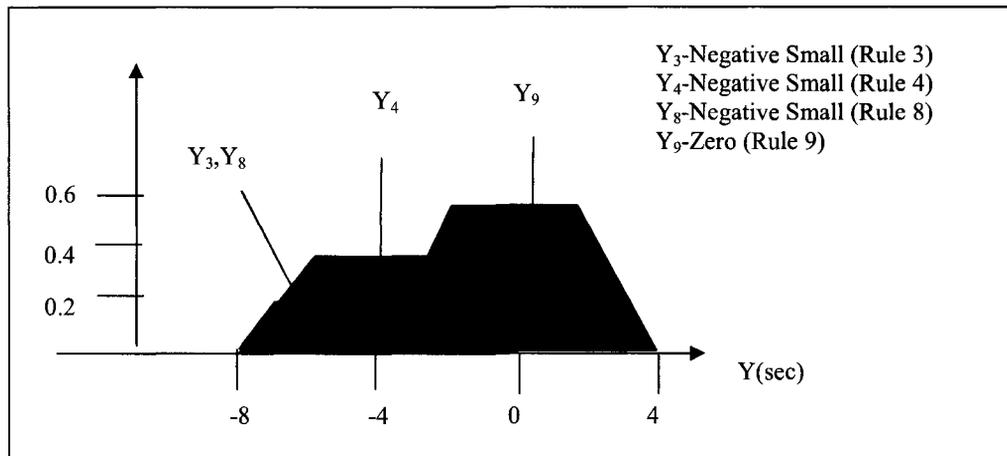


Figure 6.11: Fuzzy Output

The output is a fuzzy set with the following interpretation:

- Rule 3 suggests with strength of 0.2 that the output (change of green time) should take a value between -8 and 0 (Negative Small);
- Rule 4 suggest with strength of 0.4 that the output should be between -8 and 0 (Negative Small);
- Rule 8 suggests with strength of 0.2 that the output should be between 0 and 8 (Negative Small);
- Rule 9 suggests with strength of 0.6 that the output should be between -4 and 4 (Zero); and
- The other rules do not suggest anything.

The above suggestions are close to human thinking; however, to take control action, a crisp decision is needed. Thus the fuzzy output has to be converted into a crisp value through defuzzification. The two most frequently used methods, the Centre of Gravity (COG) and the Centre-Average methods are presented through numeric examples.

The **Centre of Gravity (COG)** method calculates the centre of gravity of the output fuzzy set relative to the vertical axes and it is calculated by using Equation 6.10 which is reproduced below.

$$r^* = \frac{\sum_i r_i \int \mu_i(Y_i)}{\sum_i \int \mu_i(Y_i)}$$

The area of the trapezoid Y_3 is:

$$\int \mu_3(Y_3) = \frac{(0 - (-8)) \times (2 - 0.2)}{2} \times 0.2 = 1.44$$

The areas of trapezoids Y_4 , Y_8 , and Y_9 are calculated similarly and the results are presented in the third column of Table 6.12. The second column contains the x coordinate of the centres of gravity of the four trapezoids.

Table 6.12: Values Used in the Centre of Gravity Defuzzification Method

Rule i	Centre of gravity v_i	Area
3	-4	1.44
4	-4	2.56
8	-4	1.44
9	0	3.36

The crisp output value is then calculated by:

$$w_{crisp} = \frac{-4 \times 1.44 - 4 \times 2.56 - 4 \times 1.44 + 0 \times 3.36}{1.44 + 2.56 + 1.44 + 3.36} = -2.47$$

The following example demonstrates the Takagi–Sugeno inference process with singleton outputs. As discussed in Section 6.3.3, in the Takagi-Sugeno process the output membership function can be represented by singleton values, which are actually constants. Assume that the value of the output Negative Small is -4 and the value of the output of the Zero is 0. The inference process is depicted in Figure 6.12. The process is similar to the Mamdani process in terms of the calculation of the firing strength of individual rules. However, the outputs are represented by crisp values instead of fuzzy sets. For example, the output of Rule 1 is negative Small defined with the crisp value of -4.

Individual rules are combined by using the Centre Average method by using Equation 6.12 which is reproduced below:

$$r^* = \frac{\sum_i a_i \mu_i(Y_i)}{\sum_i \mu_i(Y_i)}$$

It is obvious that when using this method, the shape of the output membership functions is not playing any role, thus it makes sense to use it with singleton output membership functions. The centres of gravity and membership values of the output membership functions are summarized in Table 6.13.

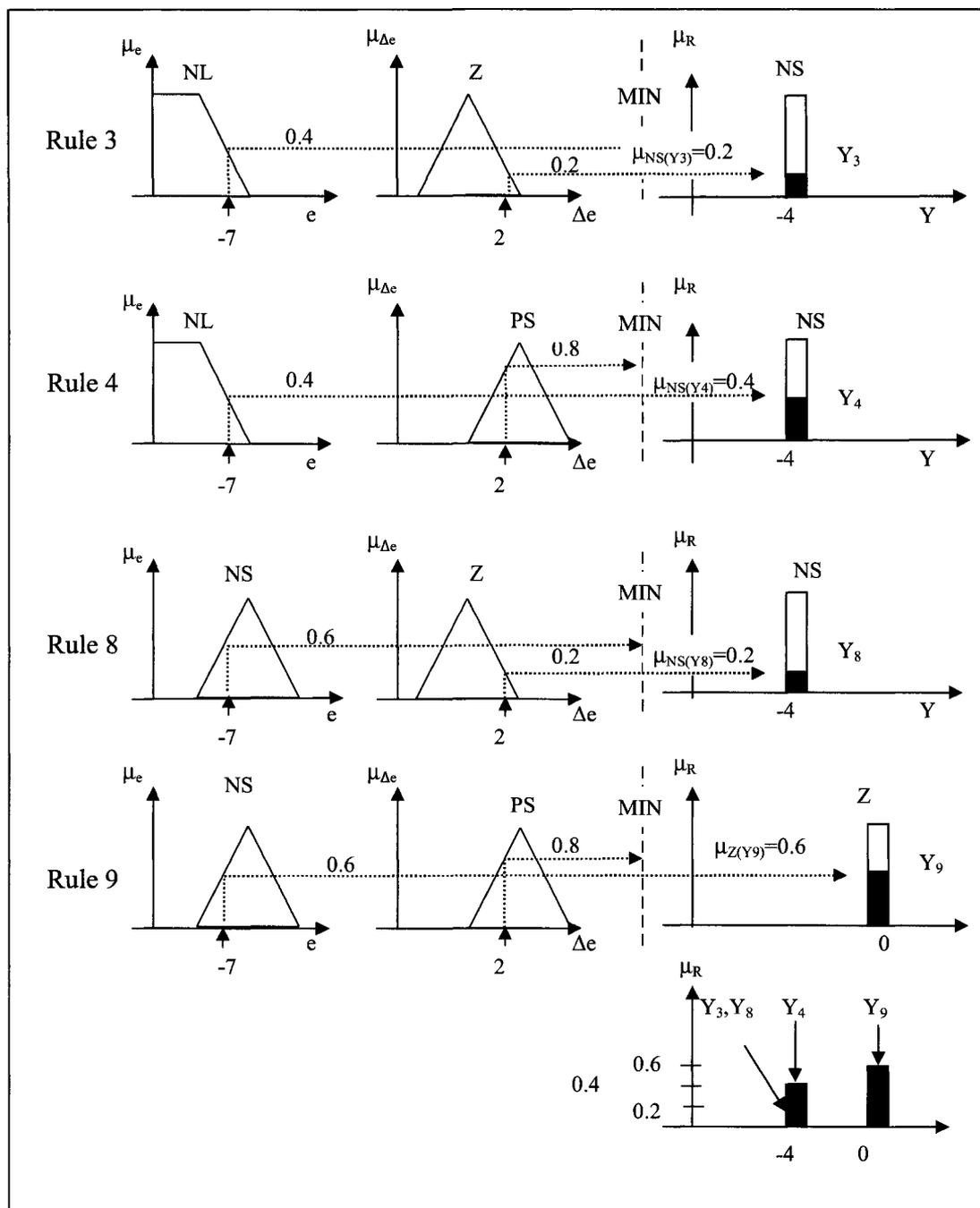


Figure 6.12: Graphical Presentation of the Takagi-Sugeno Inference Process with Singleton Outputs

Table 6.13: Values Used in the Centre Average Defuzzification Method

Rule i	Centre of gravity r_i	Membership $\mu_i(Y_i)$
3	-4	0.2
4	-4	0.4
8	-4	0.2
9	0	0.6

Then, the crisp output value is calculated by:

$$w_{crisp} = \frac{-4 \times 0.2 - 4 \times 0.4 - 4 \times 0.2 + 0 \times 0.6}{0.2 + 0.4 + 0.2 + 0.6} = -2.29$$

As expected, the two defuzzification methods provided somewhat different crisp output values.

6.5 Tuning of Fuzzy Systems

6.5.1 Tuning Objectives

Tuning automatic control systems is an essential and complex element of control system design. In traditional PID (Proportional, Integral and Derivative) control, tuning is carried out by selecting the gains for the proportional, integral, and derivative components of the control function. Even in PID control, which has a much longer development and application history than fuzzy control, tuning, in many cases, is an ad hoc process (Lewis, 1978).

In fuzzy control, tuning is normally carried out by selecting input and output membership function parameters to satisfy control objectives. The development of the rule base, the selection of the OR and AND operators, defuzzification methods are normally part of

system design and not tuning. However, various OR and AND operators, rule bases, defuzzification methods would normally require different tuning parameters.

The objective of tuning is to select controller parameters that satisfy specific control requirements. The control requirements of queue control, in the general case, are defined by:

- Accuracy and stability, and
- Responsiveness.

Accuracy means that the difference between the measured queue length Q_M and the target queue length Q_T is less than a predetermined value, ε (Equation 6.16). Stability means that the amplitude of the oscillation of the queue length is not increasing continuously (Figure 6.13).

$$\left| Q_M(k) - Q_T(k) \right| < \frac{\varepsilon}{2} \quad 6.16$$

$Q_M(k)$ - measured queue length in cycle k

$Q_T(k)$ - target queue length in cycle k

ε - permitted maximum error

Responsiveness means that the control system has to adapt within a specified time interval τ to changes to the target queue length (Equation 6.17) without significant overshoot.

For example, if the target queue length changes from $Q_{T1}=8$ vehicles in cycle k to $Q_{T2}=16$ vehicles in cycle $k+1$, the measured queue length shall reach the new target

within a specified number of cycles τ , and the overshoot should be no more than $\varepsilon/2$

(Figure 6.13)

$$T_R < \tau$$

6.17

T_R - reaction time

τ - permitted maximum reaction time

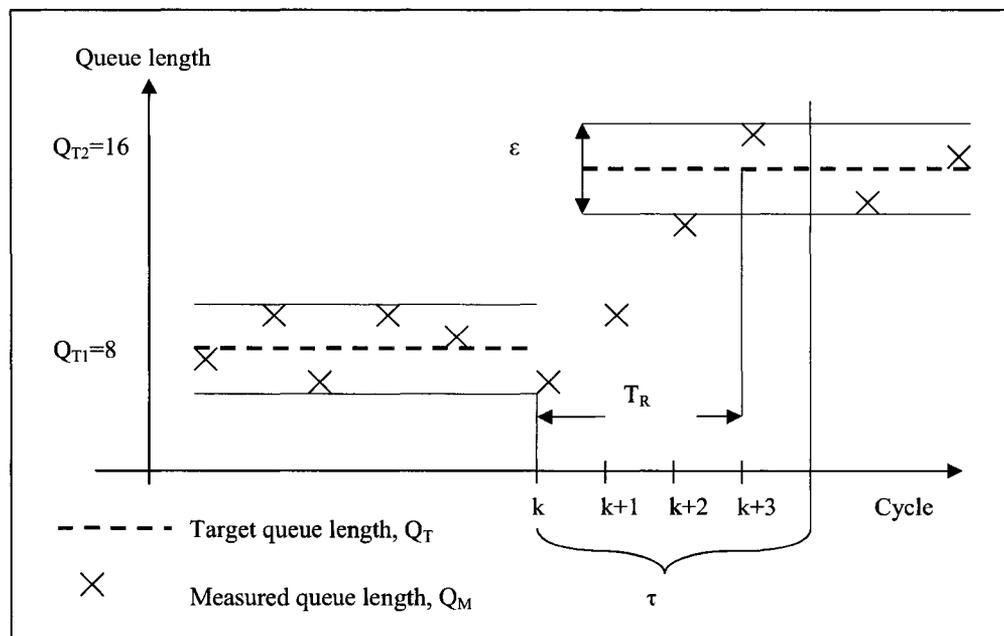


Figure 6.13: Description of Tuning Objectives

6.5.2 Tuning Techniques

Although there is no one specific procedure for the tuning of fuzzy systems that guarantees high performance control systems, there are a number of different approaches.

These include statistical techniques, neural networks, genetic algorithms, and the systematic adjustment of membership functions (Yen et al., 1998). In this thesis, two tuning techniques are utilized, the systematic adjustments of membership functions and

the neural network method. The former is described in this Chapter, and the latter is described and utilized in Chapter 10.

Although the selected tuning process does not guarantee an optimal control from the queue length control perspective, as long as the control objectives in terms of responsiveness and accuracy are met, it will guarantee that the traffic control system will achieve its objectives.

6.5.2.1 Tuning input membership functions

Tuning input membership functions is done by scaling the membership function parameters. Similarly to the normalization of input functions (Section 6.4.2), scaling changes membership functions by changing the scale of the x-axis by applying Equation 6.18 . (However, in this case input variables are not modified, only the membership functions.). Table 6.14 contains the scaled parameters using a scaling factor of $sf_e=2$ for the normalized membership functions defined in Table 6.7.

$$\{\text{scaled parameter}_i\} = \frac{\{\text{parameter}\}}{sf_i} \quad 6.18$$

sf_i - scaling factor

Table 6.14: Scaled Membership Function Parameters for the Input Variable *Error*, e

Linguistic Variable	Scaled <i>error</i> , $sf_e=2$			
	a	b	c	d
NL (-2)	-99*	-95*	-0.5	-0.25
NS (-1)	-0.5	-0.25	0	
Z (0)	-0.25	0	0.25	
PS (1)	0	0.25	0.5	
PL (2)	0.5	1.0	-95*	-99*

* - these parameters do not have to be scaled

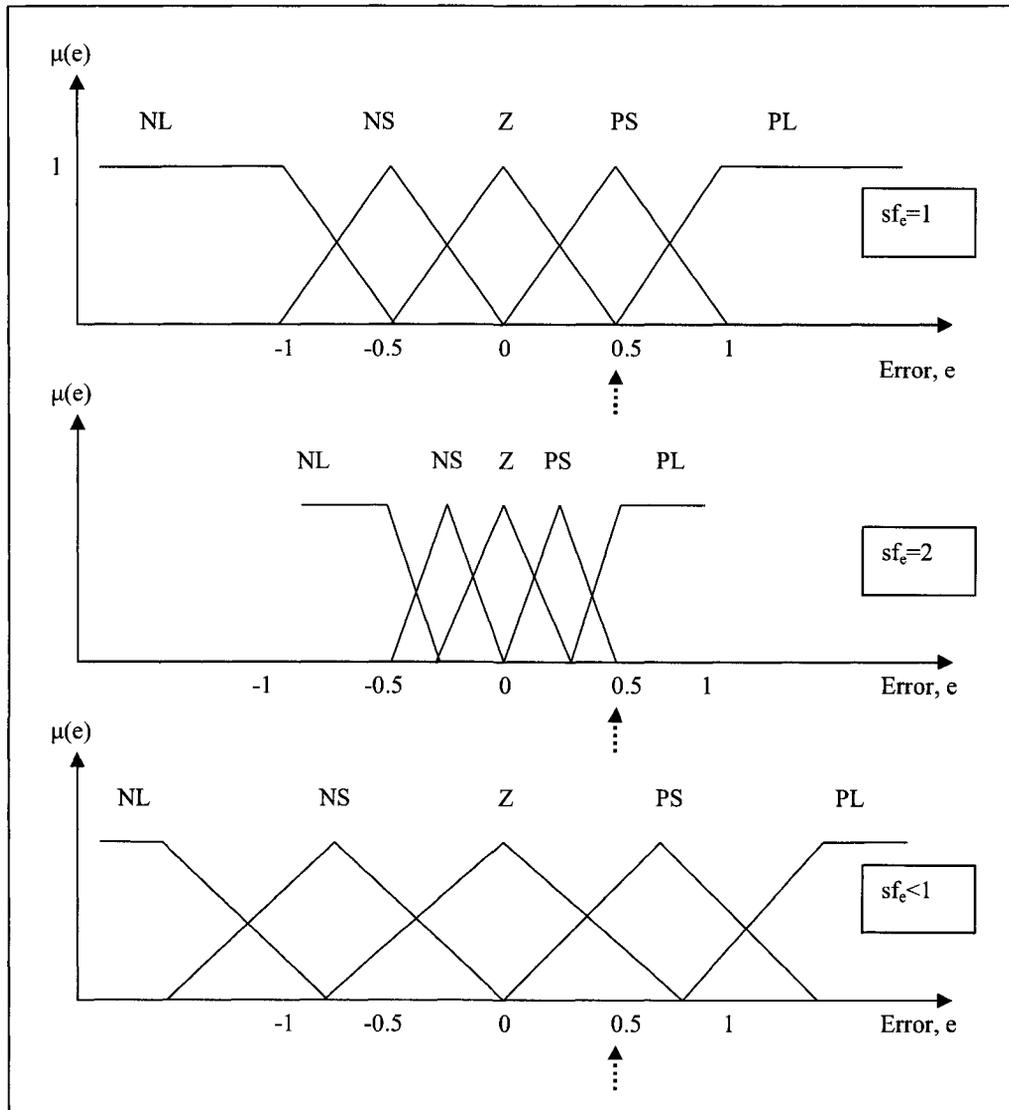


Figure 6.14: Scaling Input Membership Functions

Scaling factors above one contract membership functions, and scaling factors smaller than one expand membership functions. By changing the scaling factor of an input variable, the control system becomes more or less sensitive to the changes to that variable

(Figure 6.14). For example, an *error* of 0.5 is considered to be a Positive Small Error if no scaling factor is used ($sf_e=1$). By increasing the scaling factor ($sf_e=2$), the same *error* is considered to be Positive Large and the system will be more sensitive to *error* changes. When the scaling factor is reduced below one, the same *error* of 0.5 is considered to be somewhere between Zero and Positive Small, thus the system is less sensitive to the *error*.

6.5.2.2 Tuning output membership functions

Assume that the output membership functions are defined by linguistic variables Negative Large (NL), Negative Small (NS), Zero (Z), Positive Small (PS) and positive Large, or the corresponding numeric linguistic variables, -2, -1, 0, 1, and 2. It should be noted that both the linguistic variables (NL, NS, etc.) and the numeric linguistic values (-2, -1, etc.) are only normalized symbols and they do not represent quantities. However, the final output must be a crisp value which is obtained by defuzzification.

Tuning is designed to be part of the defuzzification process as tuning is used to attach crisp values to the linguistic output variables NL, NS, Z, PS, and PL. Two parameters are used to quantify the linguistic variables, the de-normalization factor and the scaling factor.

The de-normalization factor defines the extreme values of the outputs, i.e. the centers of NL, and PL. This is very similar to amplifying the gain of the output in a PID controller.

The scaling factor defines the centroids of all other linguistic variables which are not necessarily linearly distributed between the extremes.

The objective of scaling is to change the spacing between the output membership function centres differently at the midpoint than the membership function centres further from the midpoint (Passino et al.,1997).

Membership function centres, that are not necessarily uniformly distributed, are expressed by Equation 6.19 adapted from Passino et al.(1997) and generalized to allow scaling by any real number.

$$c_i^{sc} = r \times \text{sign}(i) \times |i|^{sf_R} \quad 6.19$$

$$r = \frac{1}{(i_{\max})^{sf_R}} \quad 6.20$$

$$i_{\max} = \text{MAX}(i) \quad 6.21$$

c_i^{sc} - scaled center of membership function i (e.g. $i=-2,-1,0,1,2$ for 5 output membership functions)

r – coefficient of scale

$\text{sign}(i)$ - sign value of i

$|i|$ - absolute value of i

sf_R - scaling factor for the output membership function centres

i_{\max} - maximum value of i

The membership function centers calculated by Equation 6.19 are always between -1 and 1 because they are normalized values. Before implementation, these values have to be de-normalized by applying 6.22. The de-normalization factor represents the maximum output value and must be determined for specific applications. For example, if the maximum output value is 10, the de-normalization factor is $dn=10$ and the output membership function centres are represented in Table 6.15.

$$c_i = c_i^{sc} \times dn \quad 6.22$$

c_i – de-normalized membership function centre

dn – de-normalization factor

Table 6.15: Scaled and de-normalized Output Membership Function Centres

MF	MF centre	$sf_R=1$	$sf_R=0.5$	$sf_R=2$
NL(-2)	c_{-2}	-10	-10	-10
NS(-1)	c_{-1}	-5	-7.07	-2.5
Z(0)	c_0	0	0	0
PS(1)	c_1	5	7.07	2.5
PL(2)	c_2	10	10	10

The impact of the scaling factor on the output membership function centres is demonstrated on the following example. Suppose that the outcome of a particular rule is to implement Positive Small change (PS or 1) and it is represented by a singleton membership function. The de-normalization factor is 10. With a scaling factor $sf_R=1$ the crisp outcome is 5, with $sf_R=0.5$ it is 7.5, and with $sf_R=2$ it is 2.5 (Figure 6.15).

If the scaling factor sf_R is larger than one, the controller is less sensitive to changes when the output is close to zero compared to when the output is further from zero. This is demonstrated by using two examples presented in Table 6.16 and Table 6.17.

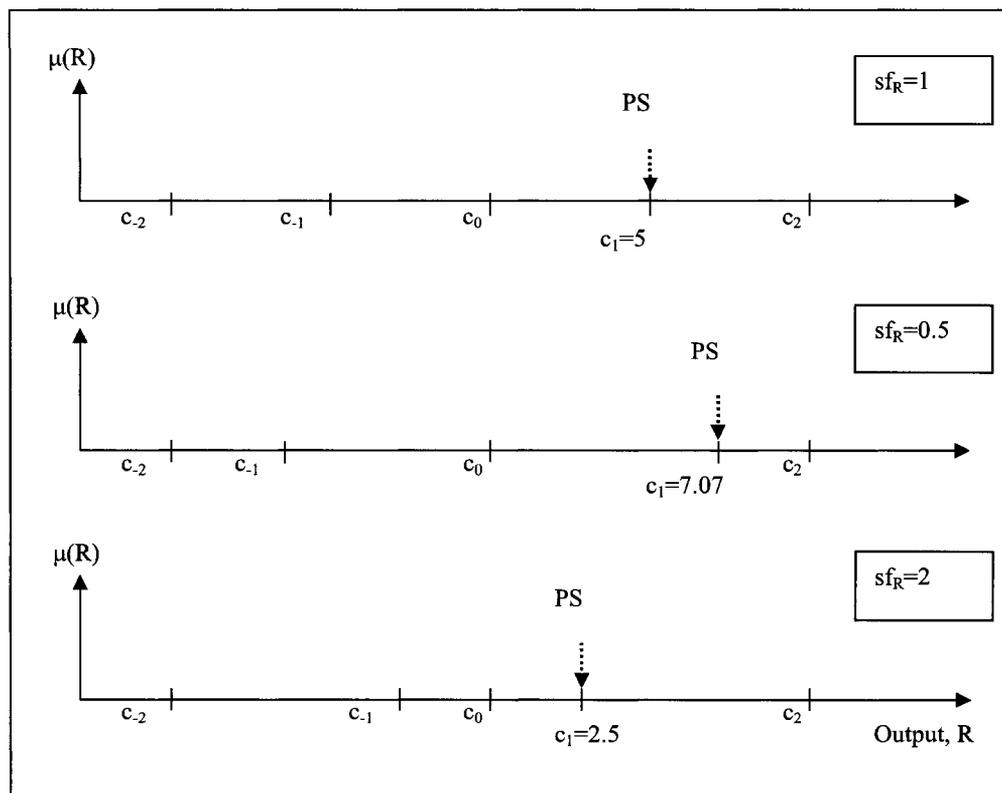


Figure 6.15: Scaled Output Membership Function Centres

The example in Table 6.16 describes sensitivity when output membership function centres c_{-1} (for NS) and c_1 (for PS) are further from output c_0 (Z) ($sf_R=0.5$). Changing the error by one unit (from -8 to -9) results an outcome change (*green adjustment*) of 0.49. However, a one unit change (from -2 to -3) will result in a much larger output change of 1.18. Thus the control system is more sensitive when the output is closer to zero.

Table 6.16: Output Membership Functions Scaled to Have Higher Sensitivity at Small Outputs and Smaller Sensitivity at Higher Outputs

$sf_R=0.5$	<i>Error</i> (Input)	<i>Green adjustment</i> (Output)	<i>Green adjustment</i> Change
Output “far” from Zero	-8	-8.19	0.49
	-9	-8.68	
Output “close” to Zero	-2	-2.36	1.18
	-3	-3.54	

Change of error=1

The second example in Table 6.17 describes sensitivity when the output membership function centres c_{-1} (for NS) and c_1 (for PS) are closer to the output c_0 (Z) ($sf_R=2$).

Changing the *error* by one unit (from -8 to -9) results an outcome change (green adjustment) of 1.25. However, a one unit change (from -2 to -3) will result in a much smaller output change of 0.39. Thus the control system is less sensitive when the output is closer to zero.

Table 6.17: Output Membership Functions Scaled to Have Lower Sensitivity at Small Outputs and Higher Sensitivity at Higher Outputs

$sf_R=2$	<i>Error</i> (Input)	<i>Green adjustment</i> (Output)	<i>Green adjustment</i> Change
Output “far” from Zero	-8	-6.67	1.25
	-9	-7.92	
Output “close” to Zero	-2	-0.83	0.39
	-3	-1.25	

Change of error=1

The difference between the output membership function scaling factors of $sf_R=0.5$ and $sf_R=2$ can be also seen by examining the control surfaces. The control surface near 0

output is much more flatter when $sf_R=2$ compared to $sf_R=0.5$ (Figure 6.16 and Figure 6.17)

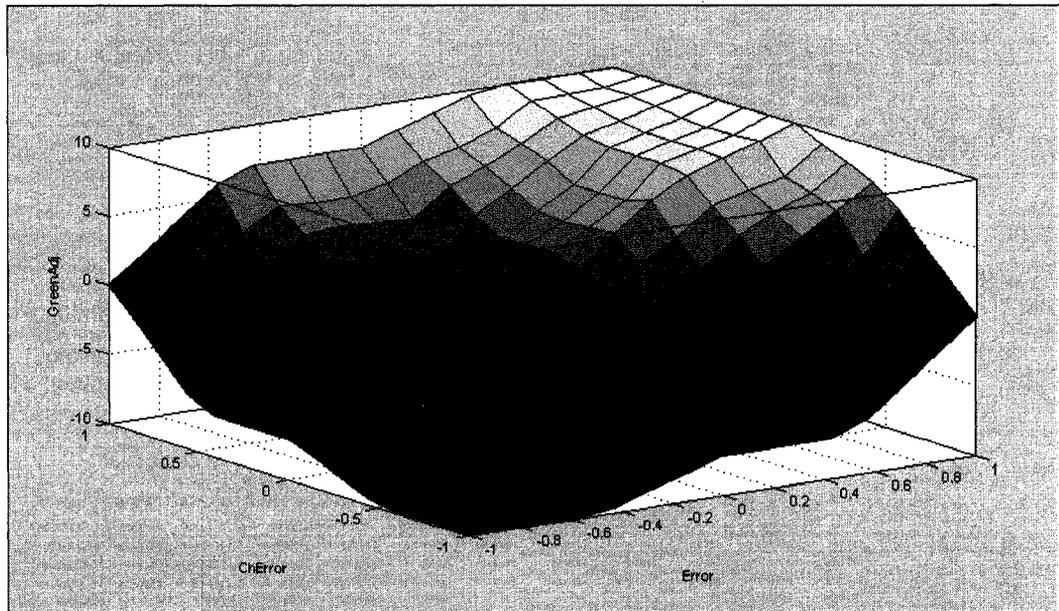


Figure 6.16: Control Surface with Output Scaling Factor $sf_R=0.5$

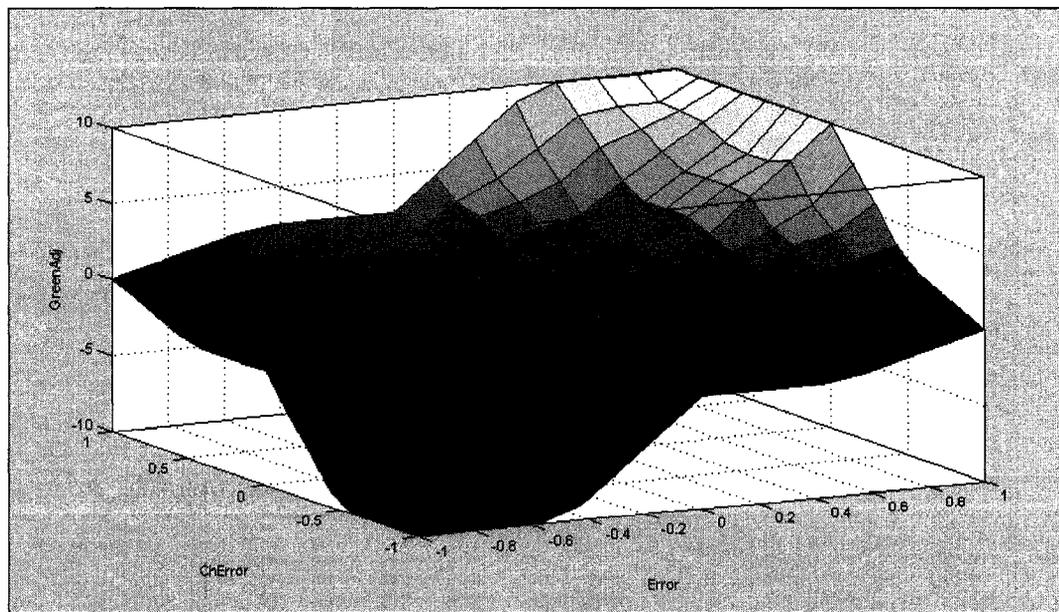


Figure 6.17: Control Surface with Output Scaling Factor $sf_R=2$

6.6 Discussion and Conclusions

Fuzzy logic provides a framework for translating human thinking into instructions useable by machines and computers. Human thinking in general is of course much more complex to be easily defined and translated to machines language, but there are system control problems where fuzzy logic has been proved to be quite effective. It is proposed that fuzzy logic could be used for queue management as well.

Fuzzy logic is based on rules which are described by linguistic terms. The fuzzy logic inference system (FIS) provides a framework to process input data (e.g. queue length) by applying the rule base and an inference mechanism to generate outputs (e.g. green signal duration).

One of the main issues in fuzzy logic control is FIS tuning. Tuning is often done by trial-and-error. However, due to the large number of tuneable parameters, the trial-and-error approach could be unreasonably labour intensive. Therefore, a somewhat systematic tuning approach is described which is based on scaling of the input parameters of the membership functions. Another tuning method is to use input–output data to train the FIS. This method is further discussed in Chapter 10.

7. DEVELOPMENT OF BIMODAL QUEUE MANAGEMENT CONCEPTS

7.1 Introduction

Bimodal queue management is based on the pragmatic assumption that a specific roadway infrastructure has a defined and limited capacity. If it is overloaded beyond its capacity, the formation of queues is unavoidable. The key objective of bimodal queue management is to design and operate the transportation infrastructure effectively even when it is overloaded.

The high level concept of multi-modal queue management is simple and can be defined as the following: “Restrict the formation of queues to locations where they cause the least damage”. Before the high level concept can be implemented, it has to be broken down to its components which are the following:

- At which locations to control queue formations, where are the capacity bottlenecks?
- How to control the formation of queues?
- How to measure the damage and benefits of queue management?

This Chapter deals with the first component. First the queue control concepts are analysed and second the analysis process for identifying queue management sites is provided. The second component is dealt with in Chapter 8 and Chapter 9. The third element has been dealt with already in Chapter 5.

7.2 How to Control the Formation of Queues

In traditional signal control, queue formation is a consequence of signal timing; it is an output. In the proposed queue control, queue formation is controlled by signal timing design; it is an input. Two distinct queue control concepts have been identified so far. The first concept applies to locations where the queue control is right at the bottleneck intersection. The second concept applies to locations where the queue control is upstream from the bottleneck intersection.

7.2.1 Queue Jump and Queue Management at the Bottleneck Intersection

The top schematic drawing in Figure 7.1 shows a typical geometric and operations environment. A queue forms at the bottleneck intersection. Assuming that the location of queue formation is acceptable, there is no need to move the queue somewhere else, thus there is no action required from that perspective. However, action is required in terms of providing roadway space that can be used for transit vehicles to pass the queue formation. The following three alternative conceptual designs were identified (Figure 7.1).

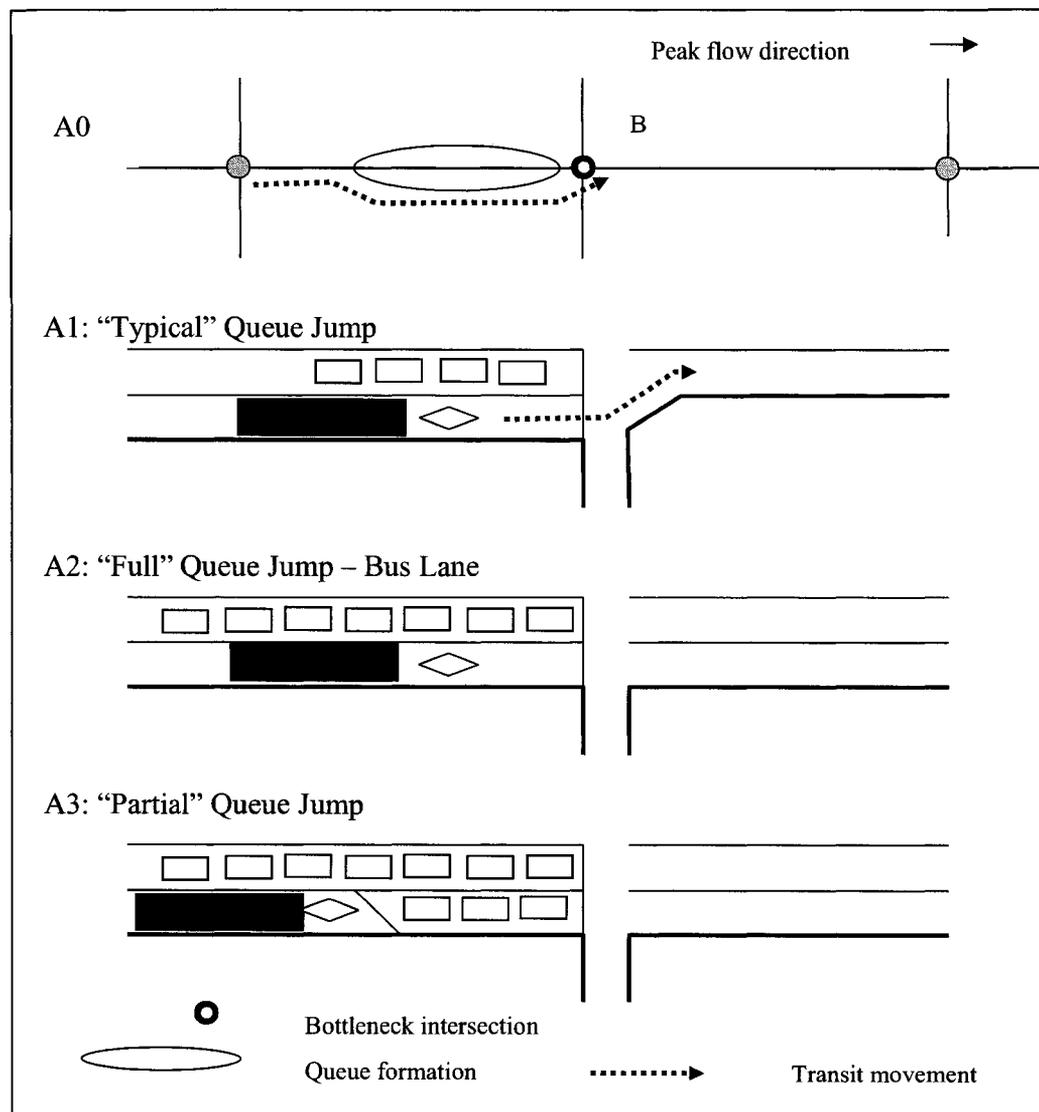


Figure 7.1: Queue Management and Queue Jump at Bottleneck Intersection

Alternative A1: "Typical" queue jump

This alternative is applied when there is less number of lanes at the far-side of the intersection than at the near-side. The extra lane at the near-side of the intersection is used as a transit queue jump lane. Implementation is relatively simple and cheap at

locations with right turning lanes without channelized right turns. If the right turn is channelized, implementation is still possible but it may be hampered by construction and utility removal costs, negative impact on pedestrian crossing times, increased minimum signal cycle time and the resulting decrease of flexibility in signal timing design. The operation of the “typical” queue jump can be enhanced by inserting a special transit signal interval into the signal cycle, but no special signal based queue management is required.

Alternative A2: “Full” queue jump

Conceptually this is a simple approach to bimodal queue management. When the length of the transit queue jump lane increases beyond the intersection approach, the measure is referred to as bus or transit lanes.

If the transit lane/queue jump lane is created by new construction, existing vehicular queues remain practically unchanged. The costs of implementation are relatively high and the trade-off between transit and car capacity is obvious. As a result, implementation requires capital investment and a progressive transit-first policy approach.

If the transit lane/queue jump lane is created by conversion, implementation costs are very low but impacts on car travel times could be significant. The reduced vehicular capacity may have significant impact on queue formation and queue blockages. Despite these disadvantages, the creation of bus lanes by converting mixed flow lanes has been implemented in many cities, mostly outside of North America. The initial negative impacts on vehicular capacity are often balanced by the long-term positive impacts on

transit service and on the social and physical environment. Implementation requires a very progressive transit-first policy approach.

The operation of transit vehicles can be further enhanced by various transit signal priority measures but there is no requirement for any special signal based queue management.

Alternative 3: “Partial” queue jump

This alternative is based on the principle that the capacity bottlenecks along transportation networks are usually at the intersections and not on the links between intersections. This practically means that links between bottleneck intersections must have certain amount of unused capacity. The “partial” queue jump uses this extra capacity to provide space for transit vehicles to pass the vehicular queues.

The greatest advantage of partial queue jumps is that they can be created by converting existing lanes while there is no or minimal impact on vehicular capacity. The conversion is carried out at midblock sections where there is extra vehicular capacity, and at bottleneck intersection approaches the full cross section is maintained for all users.

The implementation challenge of “partial” queue jumps is how to keep motorists out from the designated midblock transit lane during oversaturated conditions. Marking and signing transit lanes appropriately is not a technical or legal issue but the effectiveness of the regulatory approach can be questioned. Although lane use becomes illegal, violation is rather convenient and the benefits are tempting, thus significant violation is to be expected.

There are basically two approaches to lane designation enforcement. The first relies on traffic police presence and enforcement. The second relies on physical or other design elements which ensure that lane violation becomes obvious, illegal, and inconvenient. This approach is called “enforcement by design” in contrast to enforcement by traffic police.

There are at least two methods to implement the “enforcement by design” approach. The first method relies on a physical barrier which makes lane violation both obvious and inconvenient for the violator. The second method uses the pre-signal concept similar to the one described in the literature review (Chapter 3). The pre-signal concept requires the use of a signal control algorithm which has queue management function.

The development and application of various queue management algorithms for the “Partial” queue jump are presented in Chapter 8.

7.2.2 Queue Jump Upstream from the Bottleneck Intersection

In many situations, the provision of a transit or queue jump lane is not feasible on the link feeding the bottleneck intersection (B0 in Figure 7.2). In those circumstances it may be possible to move the transit queue jump function to the upstream intersection M (B1 in Figure 7.2). Depending on the specific geometry and operational characteristics of intersection M, a queue jump described in Section 7.2.1 and Figure 7.1 can be implemented.

If the capacity of the upstream intersection M remains higher than the capacity of the bottleneck intersection B even after the implementation of the queue jump at M (Set up B2 in Figure 7.2), signal based queue management at intersection M can be used to improve the operation of the facility. The development and application of a fuzzy logic based queue management algorithm is provided in Chapter 9.

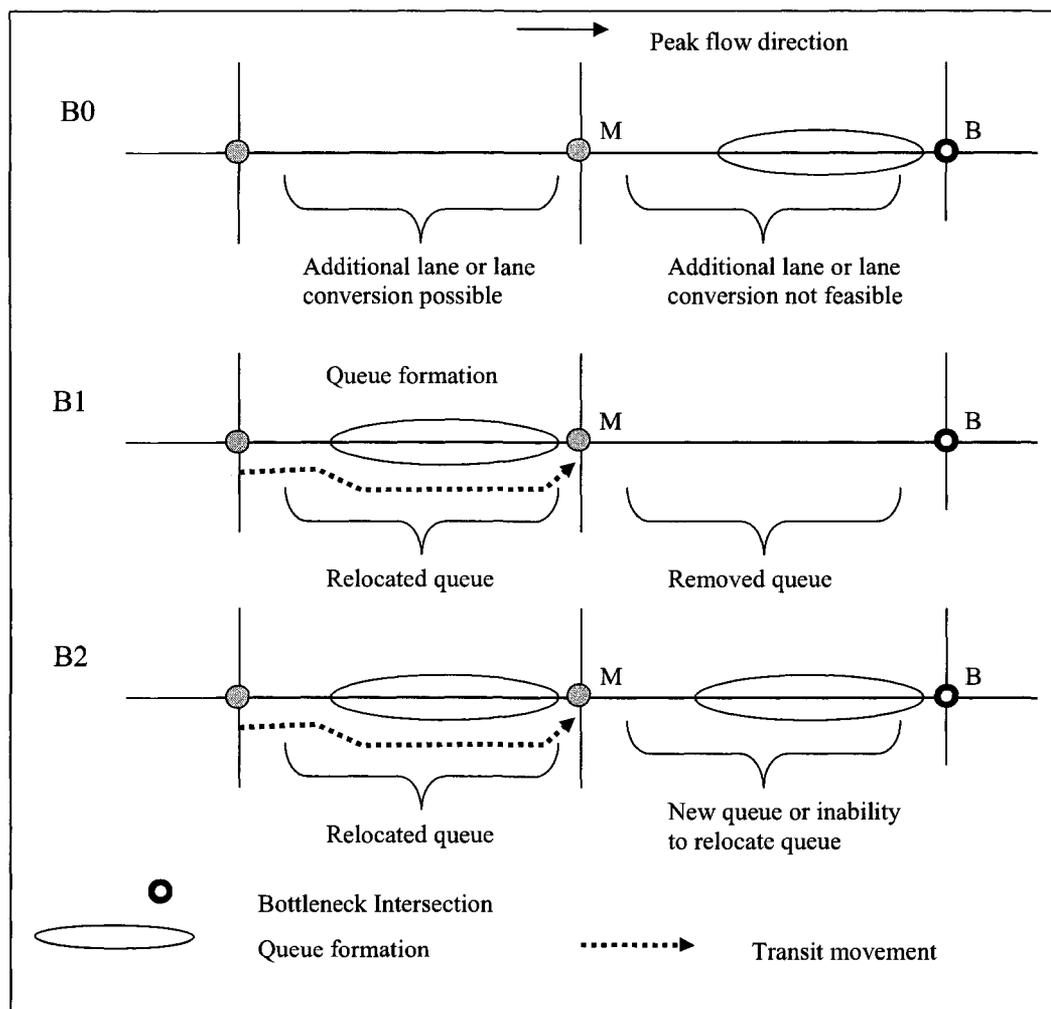


Figure 7.2: Queue Management and Queue Jump Upstream from Bottleneck Intersections

7.3 Identification of Queue Management Locations

The design of bimodal queue management measures is site specific. However, in oversaturated conditions, when secondary queues develop and impact upstream intersection operation, the identification of bottlenecks where queue management measures would be beneficial is not always straightforward. Furthermore, volume data alone, that is customarily collected, is ill suited to reflect the main bottleneck locations.

To select sites for the implementation of bimodal queue management measures an analysis of operations in both unsaturated/saturated and oversaturated conditions is required.

7.3.1 Operations Analysis in Unsaturated and Saturated Conditions

The difference between various congestion levels has been already described in detail in Chapter 3. The main difference between oversaturated conditions on one side and saturated and unsaturated conditions on the other side is that in oversaturated conditions the queues spill back to upstream intersections and impact their operation. This essentially means that in unsaturated and saturated conditions it may be possible to analyse intersections individually. However, in oversaturated conditions, the individual analysis of intersections is insufficient and could be misleading.

Along arterial roadways, signals operation is normally coordinated between adjacent intersections to provide smooth flow by minimizing the number of vehicle stops at individual intersections. The coordination is achieved by a common signal cycle length

and appropriate offset timing. The procedures for calculating signal timing parameters are well documented in transportation engineering textbooks, manuals, and research papers (Roess et al., 1998; CITE, 2008; Newell, 1989). Also, there are number of software based analysis tools developed to help with parts of the analysis of traffic operations. These include: TRANSYT, PASSER, SYNCHRO/SimTraffic, SIDRA, VISSIM, LISA, HCS, CCG Software, Integration, Paramics. Regardless of the analysis tool used, the general arterial analysis procedure is presented in Figure 7.3.

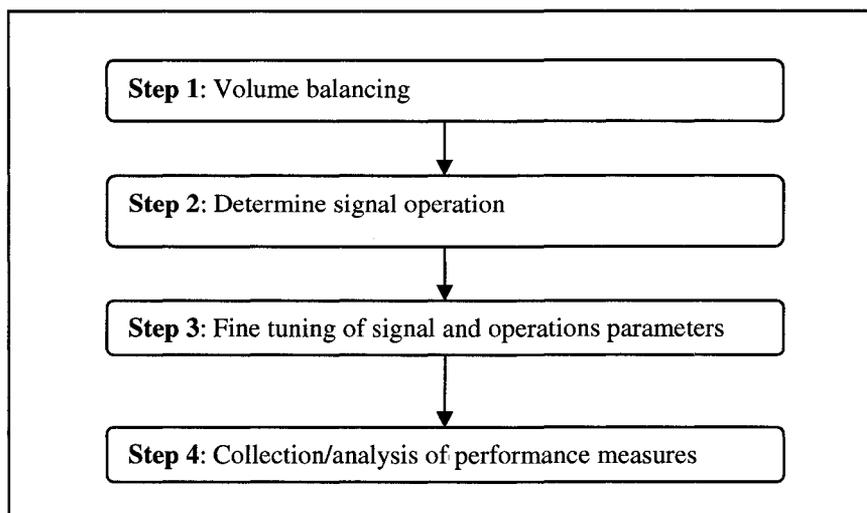


Figure 7.3: The “Traditional” Analysis Process for Unsaturated Corridors

Step 1: Volume balancing

The objective of volume balancing is to equalize input and output volumes at individual links. These volumes often differ because data collection at adjacent intersections is often carried out during different time periods. However, unless the difference between the input and output flows is over 15-20 %, the manual volume balancing is normally omitted. Some software tools such as Synchro/SimTraffic, HCS, CCG Software allow the

user to input unbalanced volumes and they base the calculations on these volumes or carry out very crude balancing by dropping extra vehicles or generating extra vehicles along the unbalanced links (SimTraffic). Other software packages, such as VISSIM have an input data structure that guarantees balanced volumes by design.

Step 2: Determining signal operation

Signal timing parameters can be collected from the appropriate authority or calculated manually or with the assistance of one of the computerized tools. Most software tools mentioned for the analysis of traffic operations are also capable to calculate signal timing parameters, but not all. For example VISSIM does not have that capability.

Step 3: Fine tuning signal and operations parameters

Fine tuning of signal timing and operations parameters is an essential element when calculated signal timing parameters are implemented in the field. If the calculated timing is for analysis purposes, fine tuning may still be beneficial. The objective of fine tuning is to check the assumptions used in the theoretical calculations, and modify them when required. The most important assumptions include the saturation flow rate, the effective length of the green time, mean travel speed and/or travel time between intersections, the number of vehicles that turn right on red, and the impacts of pedestrian movements.

Step 4: Collection and analysis of performance measures

One of the most often used performance measures is the V/C ratio indicating the sufficiency of capacity. Other performance measures include various forms of travel time/delay measurements.

7.3.2 Operations Analysis in Oversaturated Conditions

7.3.2.1 What can be Learned from Traffic Volume Data?

The analysis of operations in unsaturated and saturated conditions is usually carried out for average conditions (i.e. average traffic volumes are used during the analysis period with or without a peak hour factor). There are two compelling reasons for using average conditions in the analysis:

- it is relatively simple; and
- it is sufficient for many purposes.

Unfortunately, the same analysis process that is applicable to unsaturated and/or saturated conditions is not feasible in oversaturated conditions. The analysis of traffic operation in oversaturated conditions is much more complex for the following reasons:

- using average traffic volumes is insufficient; and
- the required geographical scope of the analysis is often unclear.

Using average traffic volumes is insufficient

The problem of using average traffic volumes in saturated or oversaturated conditions at individual intersections was discussed in Chapter 5. In summary, average conditions can be used only if demand is below capacity for most of the signal cycles during the analysis period. When demand is higher than capacity, the analysis must have a time component reflecting the duration while queues are building up and dissipating. To define this time component, one must know the distribution of arrivals within the analysis period.

The scope of analysis

In oversaturated conditions, by definition, the queues impact upstream intersection operation. Capacity bottlenecks in oversaturated conditions create chain reactions along the network and confining the analysis to a practical and manageable scope is a challenge.

The importance of finding the primary source or sources of congestion is a somewhat moot point. The location(s) of initial queue formation(s) may be far from the observed network or congestion at the initial generating points may not even exist at the time of analysis. Nevertheless, the impacts of downstream congestions introduce some difficulties in identifying bottlenecks within the observed/analysed network or road section.

The customarily collected traffic volume data usually does not contain all the information that is needed for the analysis of oversaturated corridors. If the average traffic volumes hide important information that is required for the analysis of oversaturated conditions what can one learn from the arrival volume distributions? Let's consider some examples.

The general shape of the arrival curve determined by traffic counts in short time intervals (for example 15 min) at an intersection in an oversaturated corridor is presented in Figure 7.4. The number of vehicles counted in the intervals increase, remain steady for some time, and finally decrease.

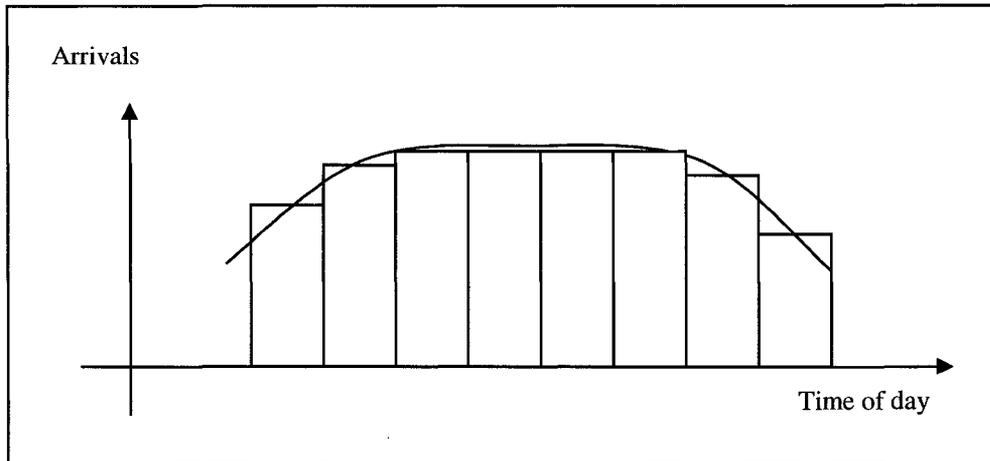


Figure 7.4: Traffic Count Distribution in Corridors with Capacity Limitations

Because the corridor is oversaturated, there is unserved demand during part of the observation period thus the horizontal part of the curve is not an indication of a constant demand. However, the horizontal middle part does not necessarily mean that the intersection approach is operating at capacity either. The only thing one can conclude from the graph is that one or more intersections in the corridor are operating at capacity.

This shape of distribution can be generated by the following three traffic conditions:

- the observed intersection is operating at capacity;
- downstream capacity is limited and the downstream queue limits the discharge rate at the observed intersection; and
- the discharge rate at the upstream intersection is at capacity.

1. The observed intersection is operating at capacity

Traffic counts are normally collected at the stop bar as vehicles enter the intersection.

When volume reaches capacity, volume counts no longer reflect demand rather they describe the capacity of the approach. As long as demand is higher or equal to capacity at

the approach, the number of vehicles entering the intersection during every signal cycle will be relatively constant. As a result, the volume count distribution curve will be horizontal (Figure 7.5). The horizontal line of the traffic count distribution represents the capacity of the intersection.

In the above circumstances, we can be certain that the horizontal part of the line represents capacity. However, when capacity is determined analytically, one cannot be sure in its exact value, thus capacity is shown as a range and not as a precise value (Figure 7.5).

The true arrival or demand distribution has a different shape, but the exact shape cannot be determined unless one counts vehicles at the end of the queue. This is often not practical because the end of the queue is constantly moving and may be impacted by upstream controlled intersections.

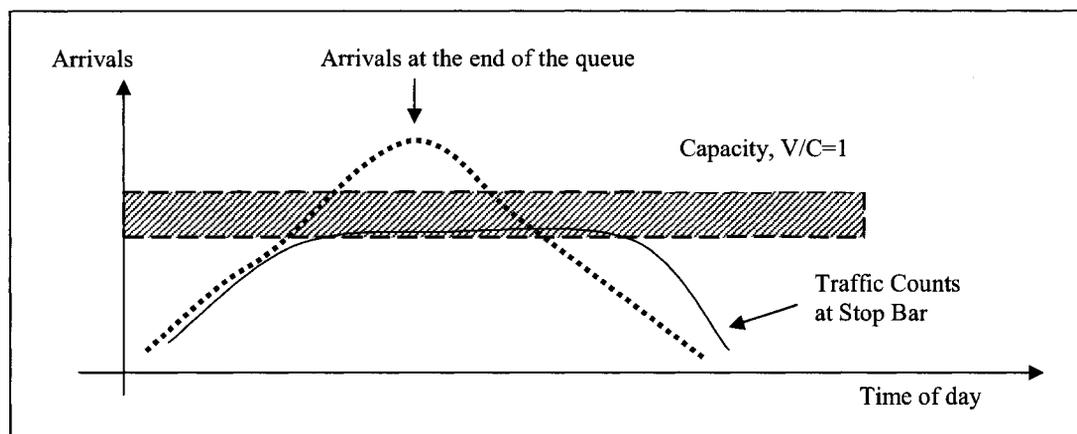


Figure 7.5: Intersection Operation at Capacity

2. Downstream capacity is limited and the downstream queue limits the discharge rate at the observed intersection

If a downstream intersection or road section reaches capacity and the end of the queue interferes with vehicle discharge at the upstream intersection where the traffic counts are taken, the resulting volume distribution will have a horizontal part even though the observed intersection has not reached its theoretical capacity (Figure 7.6).

Similarly to the previous case, the true arrival or demand profile at the observed intersection cannot be determined by collecting traffic volumes at the stop bar as vehicles enter the intersection. The true arrival distribution has a different shape; its peak may or may not reach capacity. The exact arrival distribution can be determined only with traffic counts at the end of the queue.

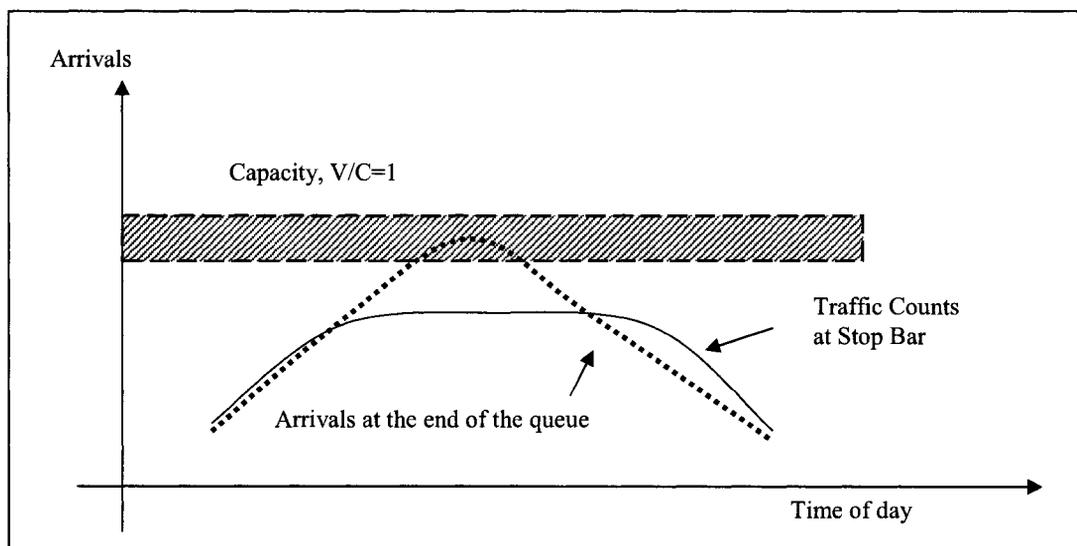


Figure 7.6: Downstream Capacity Limitations- Relationship between the Arrival Distribution, Traffic Counts Distribution, and Capacity

3. The discharge rate at the upstream intersection is at capacity

If an upstream intersection has reached capacity (discharge capacity from all movements), traffic counts at the downstream intersection approach will level off. This levelling off may occur before capacity has been reached. This is sometimes called “intersection starvation” (Figure 7.7).

The arrival distribution is somewhat meaningless in this case. If arrivals were collected at the end of the queue at the approach to the observed intersection as suggested in the previous two cases, the arrival distribution would be similar to the distribution of traffic counts. However, the true arrival profile can be described only at a location which is not impacted by upstream capacity limitations.

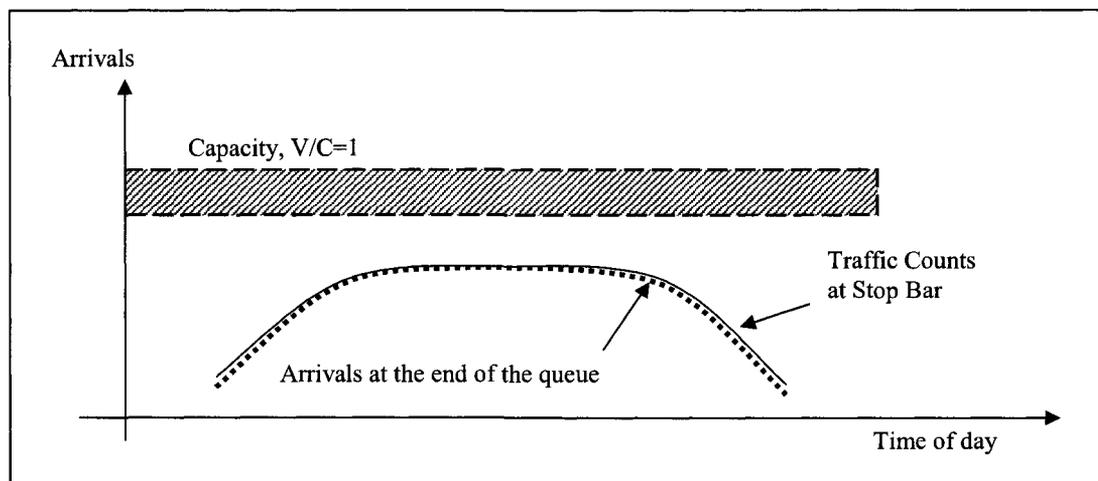


Figure 7.7: Upstream Capacity Limitations- Relationship between the Arrival Distribution, Traffic Counts Distribution, and Capacity

Based on Chapter 5 and the above analysis the following conclusions can be made:

- Analysis results in oversaturated (and saturated) condition are very sensitive to the input data. Small errors in volume data or capacity calculation generate large errors.

- Collecting precise data is problematic. The collection of volume counts at individual intersections is impacted by downstream and/or upstream bottlenecks.
- Having arrival distribution data helps to better understand the operation of the intersection, but it is still not sufficient to identify capacity bottlenecks.

To be able to analyse corridor operation in oversaturated conditions, beside the customarily available traffic counts (including arrival distribution data) and traffic signal timing data, one needs some knowledge about the roadway network even beyond the scope of the original analysis. This knowledge pertains to the locations of the primary bottlenecks or the impacts of these bottlenecks at the edges of the observed network. This point will be further explored when discussing the real-world example in the Woodroffe Avenue corridor.

7.3.2.2 Proposed Analysis Process for Oversaturated Conditions

The existence of capacity bottlenecks during oversaturated conditions requires the introduction of new steps into the operations analysis process. The proposed analysis process for oversaturated conditions, applied in this research, is presented in Figure 7.8.

The application of these steps is demonstrated on a real-world example in Chapter 8.

The proposed analysis process is based on microscopic simulation. The process is divided into two parts. The first part is based on average conditions, thus average volumes are used similarly to the “traditional” analysis in unsaturated conditions. The objective is to set-up the simulation “roughly” and to tune it for the average conditions. It is much easier

to set-up a simulation program for average conditions. Coding a microsimulation traffic network is a time consuming and error prone activity. By using constant input volume data it is much easier to ensure that the simulation is free of major coding errors and adequately tuned. The second part of the analysis process deals with the dynamic nature of oversaturation and uses input volumes based on the arrival distribution during the analysis period.

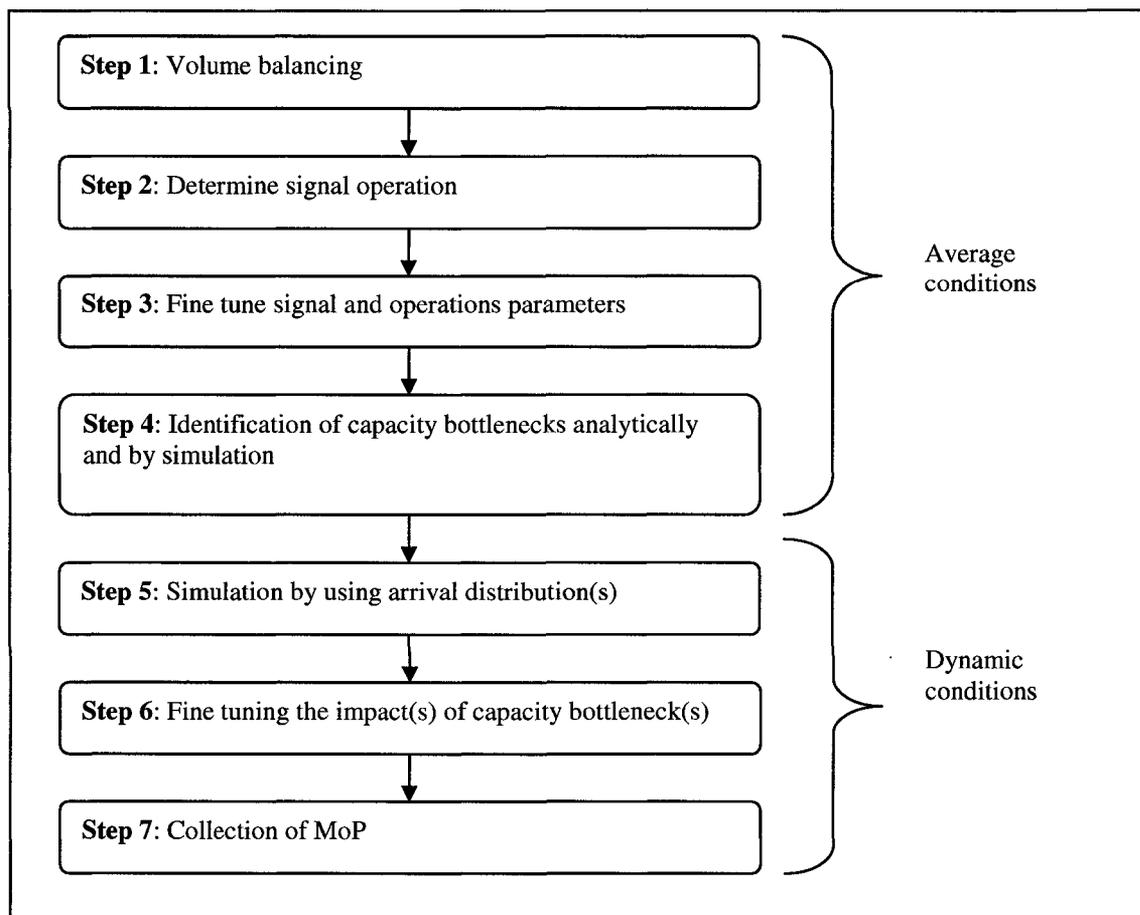


Figure 7.8: Analysis Process for Oversaturated Corridors

Step 1: Volume balancing

While volume balancing is often an optional step in unsaturated conditions, in close to saturated and oversaturated conditions it is a must. In saturated and oversaturated conditions, performance measures are extremely sensitive to input volumes. A few researchers explored pre-processing methods that aim to reconcile the inconsistencies in data collected in the field (Kikuchi et al., 1999; Kikuchi et al., 2000), but a standard procedure for volume balancing does not exist. A simple method for volume balancing was developed for the purpose of this research and it is presented in Appendix N.

Step 2: Determine signal operation

Signal timing is done the same way as for the “traditional” analysis. There are a few location specific methods in signal timing design that can alleviate the impacts of congestion. The most commonly used are reduced cycle lengths to minimize queue formations and special offsets that incorporate the existence of queues at intersection approaches. A more detailed description of these methods has been documented elsewhere (Roess et al., 1998). If these methods are appropriate, they should be included in the definition of signal operation.

Step 3: Fine tuning signal operation

The objective of fine tuning is to insure that the assumptions used in the analytical calculation are reasonably accurate. In unsaturated conditions, fine tuning of signal and traffic operations parameters is relatively straightforward because various parameters change much less from cycle to cycle. In oversaturated conditions, when queue lengths

may change from one cycle to another, driver behaviour is somewhat less predictable, collecting representative data from the field is more difficult.

Step 4: Identification of capacity bottlenecks

As discussed previously, determining the capacity of an intersection along an oversaturated corridor is somewhat problematic because of the impacts of downstream and/or upstream bottlenecks outside of the geographical scope of the analysis. As a result the V/C calculation based on average conditions may not identify the real sources of congestion but they indicate the locations of bottlenecks within the analysed section when the impacts of the other, “out of scope” locations diminish. Capacity bottlenecks may be also identified by observing simulation/animation outputs.

Step 5: Simulation by using arrival distributions

All previous steps were carried out for average conditions by using average volumes during the observation/simulation period. Using average volumes that are below capacity create steady state conditions and facilitate the simulation set up and validation.

However, operation in oversaturated conditions is not a steady state process. Average volumes are always below capacity during a long enough observation periods even if the facility is operating above capacity for part of the observation period. The proposed queue management techniques apply to those specific oversaturated time periods, therefore the input volume data must reflect oversaturated conditions.

Instead of using average volumes as input into the simulation model, the arrival distributions at the main external links have to be defined and applied. As discussed

previously, the “real” shape of the arrival distribution is not always known. But for the purpose of evaluating control techniques, the knowledge of the exact shape of the arrival profile is not essential. The shape of the distribution was estimated by using arrival distributions at intersection approaches that are unsaturated and not impacted by upstream bottlenecks. To differentiate from the “real distribution” which is unknown, it will be referred to as the “assumed distribution”.

Step 6: Fine tuning the impact of capacity bottlenecks

So far, the impacts of bottlenecks which are downstream of the observed/analysed road section were ignored. However, without these impacts, the analysis cannot be carried out.

Two main issues that need attention are:

- the quantification of impacts of downstream queues; and
- the artificial limitations of the impact of downstream queues.

The impacts of bottlenecks outside the geographical scope of analysis are simulated indirectly by modelling their upstream impacts instead of the source/mechanism of the bottleneck. This is required to keep the geographical scope of the analysis manageable.

The impact of downstream congestion can be quantified by measuring the operational capacity of the most downstream intersection within the analysis area (exiting intersection). Then the microscopic simulation has to be modified to reflect the reduction of capacity without impacting any other aspects of the simulation. Practical steps to adjust intersection capacity in microsimulation will be further explored in a real-world example in Section 8.9.

The real impacts of downstream congestion may have to be artificially reduced to prevent the propagation of queues through the observed/analysed section (geographical scope).

Queue overflows must remain within the simulated network at all times of simulation. In an oversaturated network the length of the queue is very sensitive to the input parameters.

Step 7: Data collection for evaluation purposes

The objective of data collection is to quantify the effectiveness of the proposed queue management techniques by comparing selected MoP for the existing control and the proposed control. Evaluating the effectiveness of traffic management in oversaturated conditions is more difficult than in unsaturated conditions. In oversaturated conditions the impacts, by definition, can be spread out to larger and sometimes undefined areas and the following steps are proposed to ensure that all impacts of control are included while keeping the scope of the evaluation manageable:

- the network has to be emptied from vehicles at the end of the simulation period to eliminate the impact of left over vehicles; and
- when selecting the MoP for evaluating queue management measures it is important to know the movements that are impacted by the queue management algorithm.

7.4 Summary and Conclusions

In principle there are two general locations where bimodal queue management can be implemented, at the bottleneck intersection and upstream from the bottleneck intersection. The best location for queue management is however not always obvious and an analysis of corridor operation may be required.

The analysis of oversaturated corridors is significantly different from the analysis of uncongested corridors:

- First, instead of the average traffic volumes, the volume distribution must be used to simulate operation during the period of oversaturation.
- Second, the impacts of downstream bottlenecks, which may be outside of the scope of the analysis must be included as well.
- Third, the potential consequences of congestion outside the geographical scope of analysis must be controlled.

A seven step procedure for the analysis of oversaturated conditions is proposed. The procedure is demonstrated on examples in the following two chapters.

8. BIMODAL QUEUE MANAGEMENT WITH PARTIAL QUEUE JUMPS

8.1 Introduction

This chapter contains queue management techniques developed for the partial queue jump which was introduced in the previous section. Five alternative techniques are developed, presented, and evaluated on a model example. Finally, the development and evaluation of a real-world partial queue jump is presented.

8.2 Description of Roadway Configurations

8.2.1 Physical Set-up and Operational Characteristics before Queue Management

Intersection B represents a bottleneck intersection on the network in the eastbound directions (Figure 8.1). The eastbound direction carries high car volumes and low to medium transit volumes. Signal timing at the bottleneck intersection is programmed to maximize vehicle capacity in the eastbound direction, but because demand is higher than capacity, a queue forms on the eastbound approach. The distance between the bottleneck and the first major upstream intersection is long enough to store the queue. Signal timing is set by using any of the conventional procedure such as CCG or HCM.

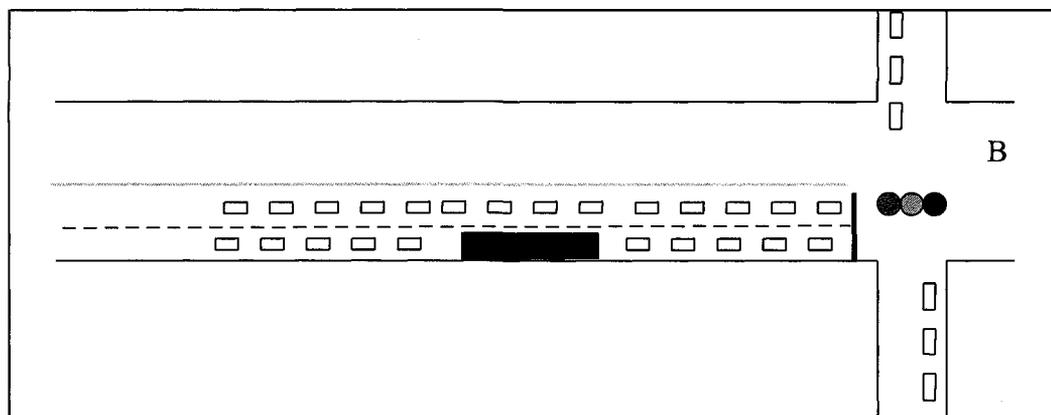


Figure 8.1: Traditional Intersection Set-Up without Queue Management

The high-level policy objective is to increase person capacity of the arterial in the long-term. However with the existing control, transit vehicles are exposed to the same congestion and unreliable travel times as car travellers and travel times are slower than car travel times, thus a significant TMS increase is not likely.

The conversion of one of the two lanes to a dedicated transit lane would theoretically provide the required increase in the person carrying capacity of the bottleneck intersection (and the arterial) but the transitions from the existing TMS to the required TMS would be too long to be socially and politically acceptable.

The above problem describes the typical traffic congestion paradox where the need for change is obvious but change is impossible (Section 4.3).

8.2.2 Physical Set-up for Queue Management

The objective of the bimodal queue management technique is to resolve the traffic congestion paradox. Change is not impossible if a control technique can be developed

during one signal cycle, the discharge rate will be maximized during the entire green phase.

In terms of reducing transit delay, the configuration depicted on Figure 8.2 provides significant benefits. The designated transit lane allows buses to overpass part of the built up queue, thus the design provides a relatively predictable intersection delay by eliminating extreme queue delays. However, this basic “partial” queue jump design has two major limitations.

The first limitation is due to the potential lack of motorists’ compliance with bus lane designations, especially at bottleneck intersections where it is feasible to achieve significant travel time gains by violating the bus lane designation for a relatively short time period.

The second limitation is due to the rigidity of design. The length of the two lane section has to be sufficient to store enough vehicles for one signal cycle to maximize capacity. To ensure that vehicular capacity is never lost due to starvation, the two lane section has to be somewhat longer than required on average. However, the longer two lane section increases transit delay. As the two lane section becomes longer, it eventually diminishes the benefits of the “partial” queue jump. Furthermore, different signal cycle lengths will have different approach capacities per cycle and different requirements for the length of the two lane cross section. As a result, the trade-offs between vehicle capacity and transit delay are unavoidable.

However, the benefits of the “Partial” queue jump outweigh these challenges, and the concept has been applied at a number of locations in Europe and more recently in Ottawa (Mucsi, 2008).

To achieve a better compliance with the transit lane use restriction and to control dynamically the number of vehicles that are stored in the restricted zone (the two lanes at the intersection approach) a design with a pre-signal as depicted on Figure 8.3 is proposed. The pre-signal operation has the following major benefits:

- capability to control dynamically the number of vehicles that enter into the restricted zone (space between the bottleneck and pre-signals);
- reduce/eliminate bus lane violation by motorists; and
- alleviates potential bus merging difficulties.

The concept of queue management with pre-signal and transit queue jump is not totally new. The literature research already provided some background information on such implementations. What is new in this research are the control algorithms developed for the operation of the pre-signal.

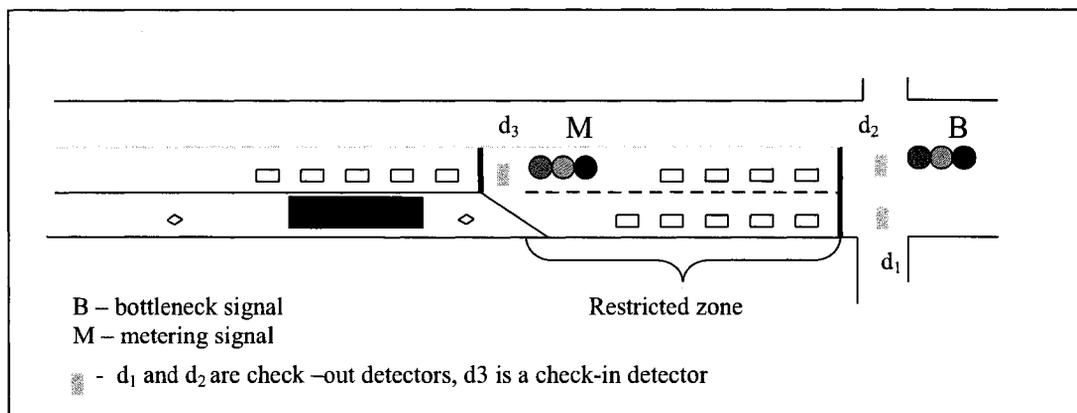


Figure 8.3: Queue Management with Pre-Signal

8.2.3 Relevant Design Input Parameters

To maximize/maintain the vehicular capacity at the bottleneck intersection, the storage capacity of the restricted zone should be able to accommodate all vehicles that can clear the intersection during one signal cycle. The basic calculations for the required parameters are provided below.

Queuing space in the restricted zone

The queuing space, expressed in vehicles, can be calculated by using 8.1.

$$L_Q = \frac{\sum_{i=1}^K L_i}{L_v + L_g} \quad 8.1$$

L_Q - available queuing space in the restricted zone (vehicles)

L_i - length of lane i that can be used for queuing (m)

K – total number of lanes in the restricted zone

L_v - average vehicle length (m)

L_g is the average gap between vehicles (m)

Green stage capacity at the bottleneck intersection

The green stage capacity (per cycle) at the bottleneck intersection B can be calculated by using 8.2:

$$cap^B = g^B \times \frac{s^B}{3600} \times n \quad 8.2$$

cap^B - cycle capacity for the movement (veh/cycle)

g^B - length of the green signal for the movement (sec)

s^B - average saturation flow rate of the movement (veh/hr)

n - number of lanes

Number of vehicles in the restricted zone

The number of vehicles that are allowed into the restricted zone for each signal cycle should be somewhat larger than the capacity of the cycle to minimize the possibility of intersection starvation (unutilized green time at the bottleneck signal).

$$Q_T = cap^B + \delta \quad 8.3$$

Q_T - the target queue length (veh)

δ - queue error margin (veh)

The error margin ensures that the target queue Q_T is larger or equal to the phase capacity at B even when the s^B and/or g^B in a particular cycle are larger than the average values.

This prevents intersection starvation.

Measured queue length (or number of vehicles in detection zone – NVDZ)

The queue length in this application is defined as the number of vehicles waiting for the green signal. The queue length (or NVDZ when there is no stationary queue) is an input into the control algorithm and it is determined by counting vehicles as they enter and exit the restricted zone (8.4) (or by estimating the number of vehicles in the restricted zone based on detector occupancy data, Chapter 10).

$$Q_M = N_{d3} - (N_{d1} + N_{d2}) \quad 8.4$$

Q_M - measured queue length (veh)

N_{d3} - cumulative detector count at the check-in location

N_{d1} and N_{d2} - cumulative detector counts at the check-out location

In terms of sampling of the queue length, two methods are used. Some queue management techniques require continuous sampling, while others sample at a specific time during the signal cycle.

8.3 Queue Management with Partial Queue Jump and no Pre-Signal (QMnoPS)

The queue management concept without a pre-signal (depicted in Figure 8.2) is attractive because of its simplicity. In terms of the physical infrastructure, only the lane designation has to be changed with appropriate signage and no pre-signal installation is required. The key for the successful operation of this measure is in the ability to minimize bus lane violation. Driver compliance can be enforced through physical measures that prevent cars

from driving in the bus lane or by strict legislative measures that are enforced. An example of the physical measure is presented in Figure 8.27.

The provision of a physical barrier to enforce a partial queue jump is not feasible when buses travel straight through or turn right at the intersection. Without “enforcement by design” continuous police enforcement would be required. However, the assumption of regular and continuous police enforcement is seldom realistic, and queue management designs that are self-enforced are expected to be more effective. The following designs all rely on the installation of a pre-signal. The difference between the designs is in the signal control algorithms applied for queue management at the pre-signal.

8.4 Partial Queue Jump with “Park View” Type Control

A queue management algorithm used in London, UK has been introduced in the literature review (Chapter 3). In this section, the original algorithm is analysed and then modified for general queue management application.

8.4.1 Algorithm Description and Analysis

The geometric set-up is shown in Figure 8.4. The original queue management algorithm as provided by Oakes et al. (1995) is shown in Figure 8.5.

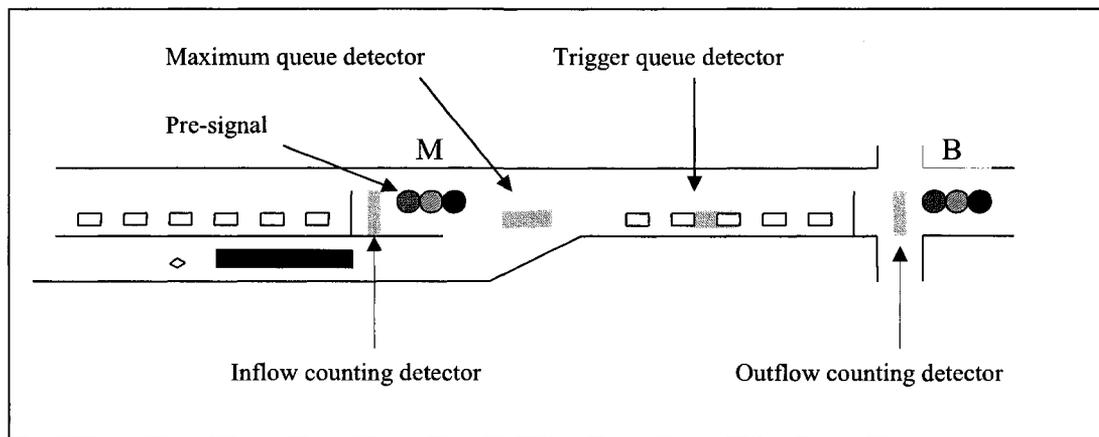


Figure 8.4: Geometric Set-Up of the Original “Park-View” Pre-Signal

The operation of the algorithm, as described by the authors, contains the following main elements:

Queue relocation triggering criteria:

This is described by block 1 Figure 8.5. It is used to turn the queue relocation mode on. There are two alternative approaches, through a master time clock (if queue formation is recurrent and predictable), or by detecting queue formation with queue detectors. The trigger queue detector is shown in Figure 8.4.

Green time adjustments to maintain queue length: This is described by blocks 2, 3, 3.1, and 3.2. The number of vehicles leaving both the upstream M (Metering) and the downstream B (Bottleneck) intersections are counted and the difference Δ is calculated. If the difference is negative, the green time g at the metering signal is increased by a pre-set value, which is set to two seconds. If the difference is positive, the green time g is reduced by two seconds.

Adjustment for too long queues: This is described by block 4 and 4.1. The green time adjustment described above will not prevent gradual queue increase or decrease. If the queue length gradually increases and reaches Q_{max} , the green time is reduced by four seconds.

Queue relocation exit criteria: This is described by block 5. When the pre-signal green time reaches a maximum length, it is assumed that there is no need for further flow restrictions at the metering intersection and the queue relocation mode can end.

The objective of the analysis of the original algorithm is to determine whether it could be used as a general concept for developing queue management techniques. The analysis revealed that the original algorithm does not prevent intersection starvation. The instructions in Block 4 (Figure 8.5) provide for the circumstance when the queue length is too long. A similar block is also needed (but it is missing) for the situation when the queue in the restricted zone is too short and intersection starvation occurs even though there is un-served demand waiting at the pre-signal.

Therefore, the Oakes' method was modified to include the missing functionality and generalized to be tuneable for different situations. The general geometric set-up is depicted in Figure 8.6 where the required parameters are defined. The new algorithm is depicted on Figure 8.7. Blocks 3, 3.1, 5, and 5.1 are new elements and the incremental green times are shown as variables Δ_1 and Δ_2 .

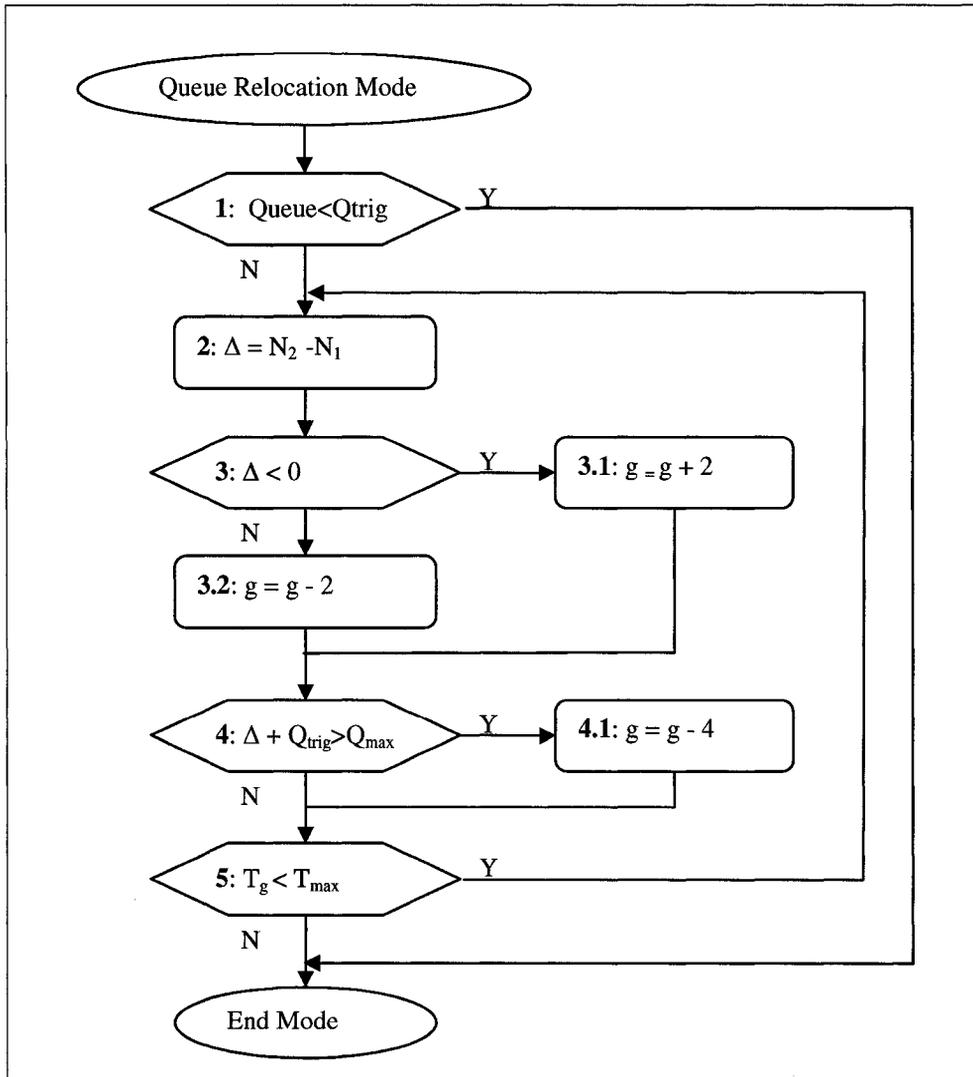


Figure 8.5: The Original Pre-Signal Algorithm by Oakes et al., 1995

The algorithm does not have single target queue length. The objective is to maintain the queue length between a minimum (Q_{trig}) and a maximum (Q_{max}) values by equalizing the inbound (N_2) and outbound (N_1) flows through minor timing changes expressed as Δ_1 (Figure 8.6). When despite these minor changes the queue length drifts and becomes shorter or longer than the minimum (Q_{trig}) or maximum set queues (Q_{max}), the green time is changed by a bigger increment, expressed as Δ_2 .

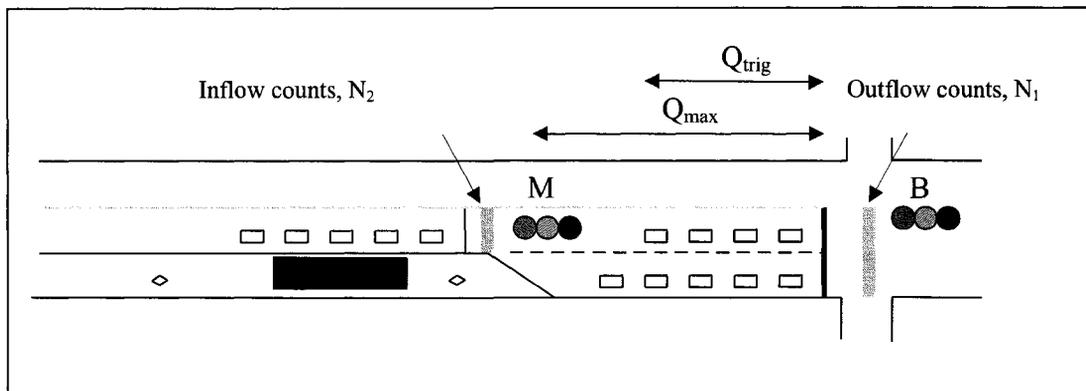


Figure 8.6: Parameters of the "Park View" Type Queue Management

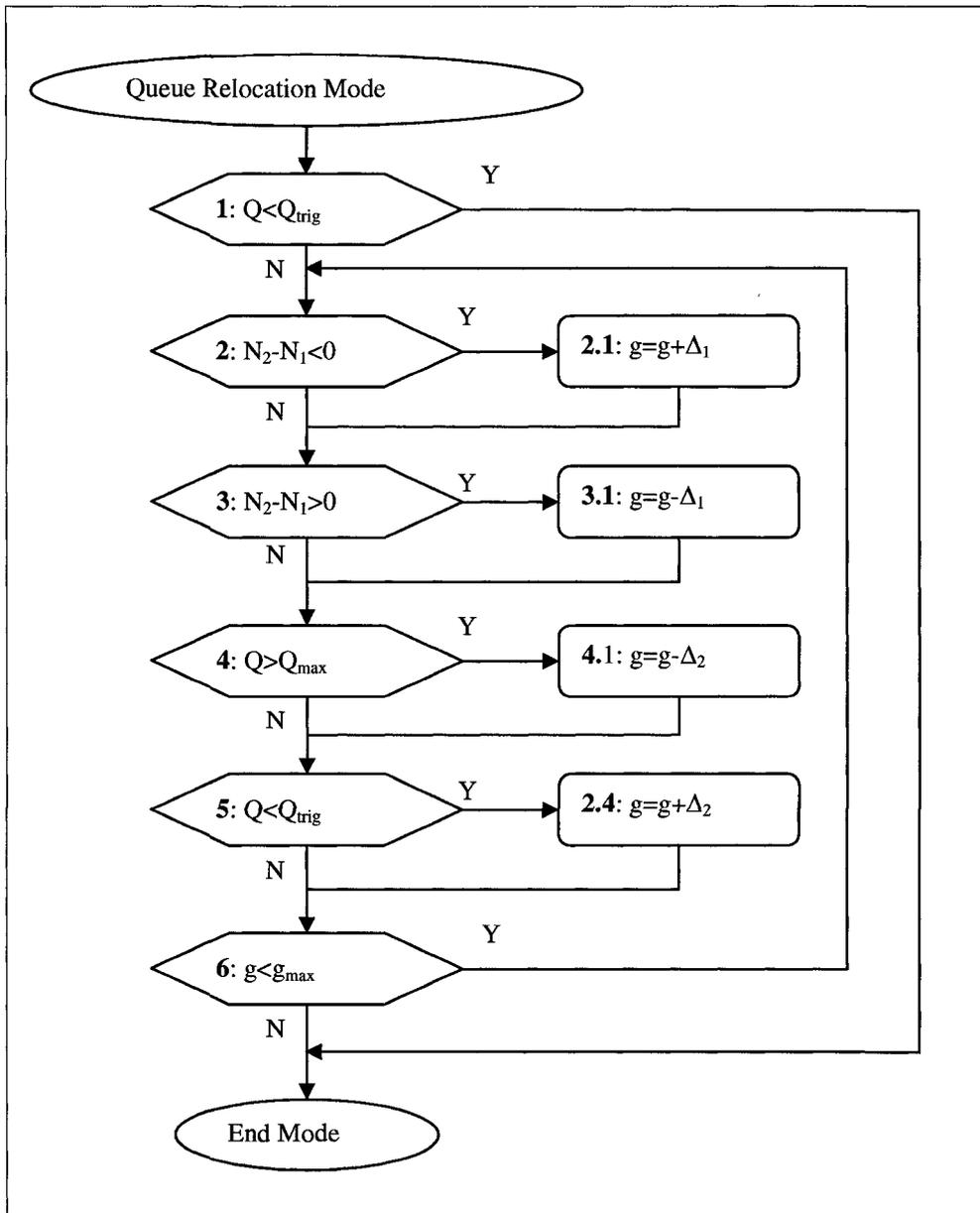


Figure 8.7: Generalized and Enhanced Queue Relocation Algorithm Originally Described by Oakes et al. (1995)

8.4.2 Algorithm Tuning

Before the algorithm can be evaluated and compared to other algorithms, it has to be tuned for the specific location and operational objectives. The tuning has primary and secondary objectives.

The primary objectives include output accuracy and system responsiveness. In terms of system accuracy, the requirement is to have a queue length within a pre-specified range which was arbitrarily selected as +2 and -2 vehicles from the set target queue length.

In terms of responsiveness, the objective is to achieve the required accuracy within the next cycle after the target queue length has changed. The applied accuracy criteria of one cycle, serves only as examples used for comparing various algorithms. Responsiveness is not required in every queue management application, it is only required in transit signal priority applications where the target queue length may change from one cycle to another. These applications are not discussed in this thesis, nevertheless, the responsiveness criteria is included to ensure a comprehensive control algorithm evaluation.

Table 8.1: Queue Control Objectives

Queue control objective	Measurement
Accuracy, maximum overshoot	$\epsilon/2 \leq 2$
Responsiveness	Within one cycle: $\tau \leq 1$
Minimize queue length variation	RMSE
Minimized green adjustment variation	$\sigma(\text{green adjustment})$

The secondary objectives include the minimization of variation of the measured queue and the minimization of the variation of the green time at the pre-signal (metering intersection).

The variation of the measured queue is expressed as the Root Mean Square of Error (RMSE) expressed by Equation 8.5. The variation of the green time at the pre-signal is expressed as the standard deviation of the green time or green time adjustments. Since the “Park View” type algorithm does not have a single value for the target queue length, the average of the minimum (Q_{trig}) and maximum (Q_{max}) queue lengths is used as the target queue (Q_T).

The queue management algorithm that controls queue lengths more tightly (lower RMSE) can be implemented at more locations because it allows queue control within less roadway space. Queue management algorithms that operate with lower green time variation are preferable from the driver expectation perspective.

$$RMSE = \sqrt{\frac{\sum_{k=1}^N (Q_M(k) - Q_T(k))^2}{N}} \quad 8.5$$

RMSE - root mean square error

$Q_M(k)$ - measured queue length in cycle k

$Q_T(k)$ - target queue length for cycle k

N - number of data points

The tuneable parameters are Q_{trig} , Q_{max} , Δ_1 , and Δ_2 . Tuning has been carried out by trial and error. Through experimentation, the following three step procedure has been followed:

- select minimum (Q_{trig}) and maximum (Q_{max}) lengths to achieve/improve accuracy;
- select Δ_1 that achieves/improves accuracy; and
- select Δ_2 that achieves/improves responsiveness.

The above procedure will find one of the local optimums but not necessarily the global optimum. However, for the control system to achieve its operational objectives, the queue management sub-system does not necessarily have to operate at its global optimum. The local optimum that maintains maximum vehicular approach capacity while it keeps the queue within a specific range is sufficient.

The physical set up is depicted on Figure 8.6, the signal timing parameters at the bottleneck intersection are summarized in Appendix F. Input volumes are set to oversaturate the intersection during the entire simulation period. During the tuning process, the bus volumes are set to zero. To test system responsiveness, the target queue length is set to alternate between 8 and 14 vehicles every 20 cycles. Total simulation time is 120 cycles or 14400 seconds.

8.4.3 Control Performance of the “Park View” Type Algorithm

Sample results of individual runs with different tuning parameters are provided in Appendix H. The final results, tuning parameters and performance measures are shown in Figure 8.8 and Table 8.2.

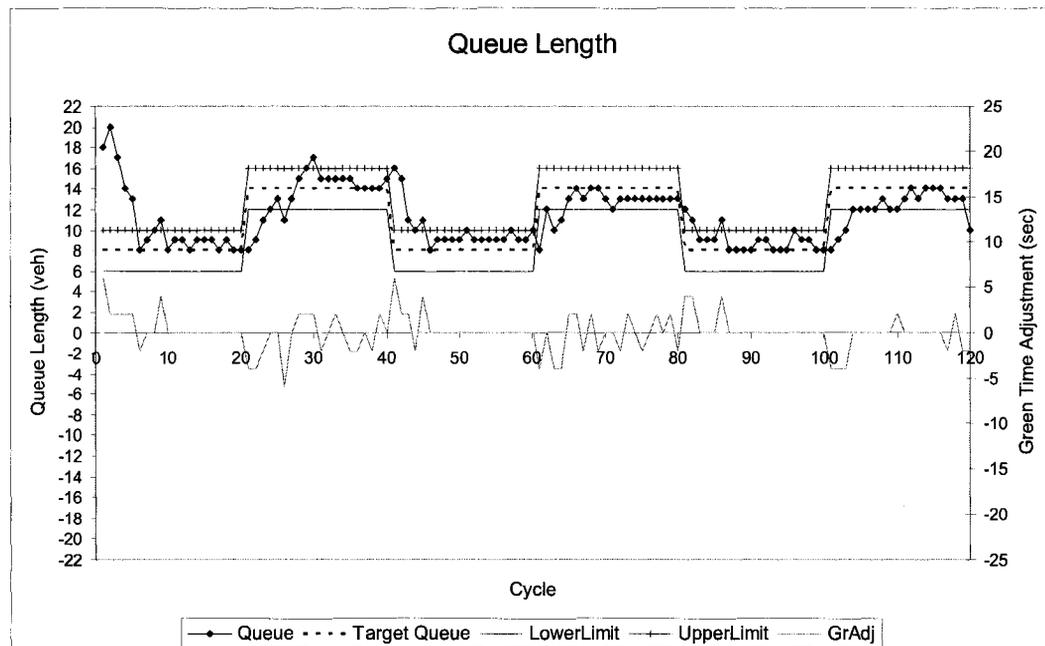


Figure 8.8: “Park View” Type Queue Control Algorithm Output with Tuned Parameters

In terms of accuracy, the “Park View” type queue control is satisfactory. It is capable of maintaining the queue length within the required error margin (± 2 vehicles from the target queue length, Q_T). In terms of responsiveness, however, the algorithm fails. For example, when Q_T changes from 8 vehicles to 14 vehicles, it takes 4 signal cycles to adjust the green time to generate the new queue length.

Table 8.2: Tuned Parameters and Control Performance for the “Park View” Type Control

Limits for increment Δ_2^*	GreenAdj. Increments	Accuracy # of points outside $Q_T \pm \tau$	Responsiveness Reaction time delay	RMSE	$\sigma(\text{GrAdj})$
$Q_T \pm 2$	$\Delta_1=2, \Delta_2=4$	22	4,4,2,2,4	2.17	1.846

*Lower limit for increment Δ_2 is $Q_{\text{trig}}=Q_T - 2$ and upper limit for increment Δ_2 is $Q_{\text{max}}=Q_T + 2$

8.5 Partial Queue Jump with “On-Off” Type Control

8.5.1 Algorithm Description

This algorithm is based on the continuous monitoring of queue lengths in the restricted zone and terminating vehicle discharge at signal M when the target queue length Q_T is reached. The pre-signal remains red until the start of the next cycle. This control represents one of the simplest approaches in control engineering.

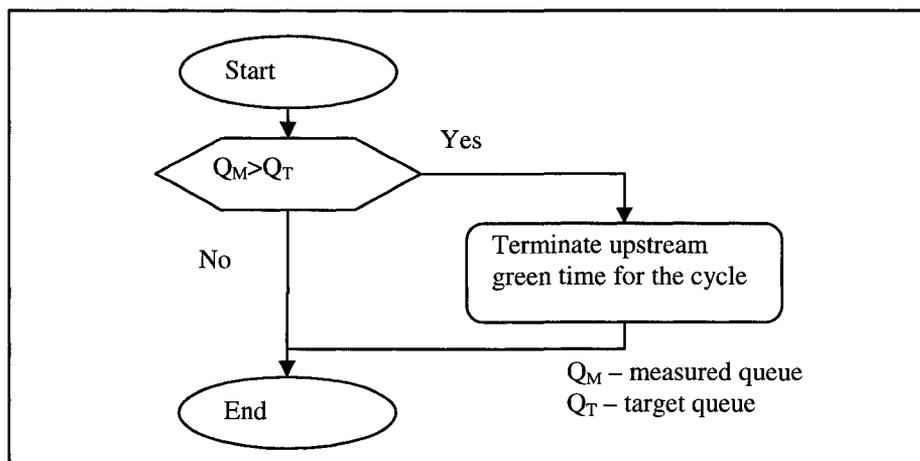


Figure 8.9: “On-Off” Type Algorithm

8.5.2 Control Performance of the “On-Off” Type Algorithm

Roadway configuration, signal timing parameters at intersection B, the approach volume, the target queue length Q_T , are identical to the example discussed in Section 8.4. The performance of the control algorithm is shown in Figure 8.10 and Table 8.3.

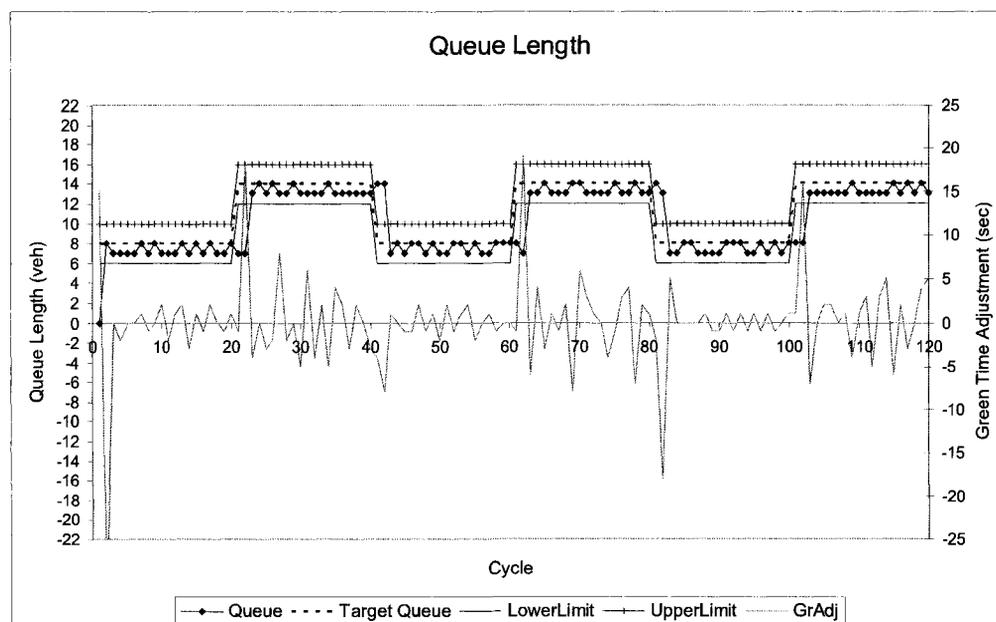


Figure 8.10: “On-Off” Type Queue Control Algorithm Output

In terms of accuracy, the “On-Off” type control is performing quite well, it maintains the target queue length with very little error. In terms of responsiveness, however, it fails the set objective, it cannot adjust to new target queue lengths within one cycle, and it needs two cycles.

Table 8.3: Control Performance for the “On-Off” Type Control

Accuracy # of points outside $Q_T \pm \tau$	Responsiveness Reaction time delay	RMSE	$\sigma(\text{GrAdj})$
10	2,2,2,2,2	2.029	4.447

8.6 Partial Queue Jump with Model Based Control

8.6.1 Model description

The Model based control uses mathematical formulas to describe traffic flow at the bottleneck intersection B and metering intersection M in cycle k-1 to predict the required green time at signal M in cycle k+1. The required data, collected continuously includes the queue length measurement in the restricted zone, the discharge volume at intersection B, and the saturation flow at signal M (Figure 8.11).

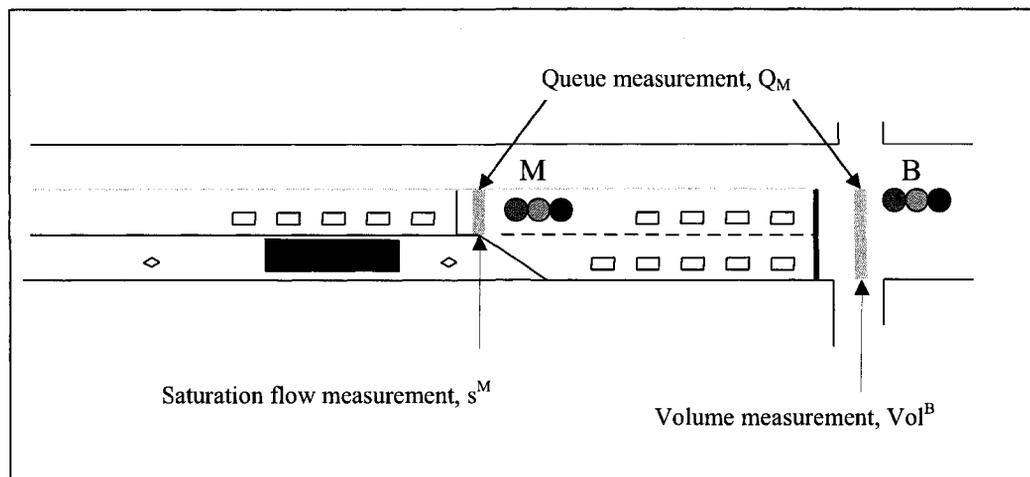


Figure 8.11: Required Data for the Model Based Queue Control

Vehicle capacity is determined by Equation 8.6. This capacity can be utilized only if there is sufficient supply of vehicles from the upstream metering intersection M.

$$C^B = s^B \times \frac{g}{cyc} \times n \quad 8.6$$

C^B – vehicle capacity at the bottleneck intersection (EB movement) (veh/hr)

s^B – saturation flow rate measured continuously for every cycle (veh/hr/lane)

g – duration of the green signal (sec)

cyc – cycle length (sec)

n – number of lanes

The upstream metering intersection has to provide sufficient green time to prevent the starvation (and capacity) loss of the downstream bottleneck intersection, however, to minimize transit delay, it must not provide too much green time either. If too much green time is provided, the space between the bottleneck and metering intersections will be filled up and there will be no space for transit vehicles, thus the potential to minimize transit delay will be reduced, although not eliminated.

The following steps are required to calculate the green time at signal M for cycle k.

The number of vehicles that have to be discharged at signal M during cycle k is determined by Equation 8.7.

$$Vol^M(k) = Q_T(k) - [Q(k-1) - Vol^B(k-1)] \quad 8.7$$

$Vol^M(k)$ – volume discharged during cycle k at signal M

$Q_T(k)$ – target queue length for cycle k

$Q(k-1)$ – measured queue length in cycle k-1

$Vol^B(k-1)$ - measured discharge volume at intersection B during cycle k-1

The green time required to discharge the volume described by Vol^M at signal M is calculated by using Equation 8.8 and 8.9.

$$Vol^M(k) = g(k) \times \frac{s^M(k-1)}{3600} \quad 8.8$$

$$g^M(k) = [Vol^B(k-1) + Q_T(k) - Q(k-1)] \times \frac{3600}{s^M(k-1)} \quad 8.9$$

$g^M(k)$ - green time at intersection M for cycle k (sec)

$s^M(k-1)$ – saturation flow rate during the green phase in cycle k-1 (veh/hr)

The queue management algorithm calculates the green at intersection M (pre-signal) by using Equation 8.9. The input data includes the volume at intersection B during the previous cycle, the target queue length, the queue length in the previous cycle, and the saturation flow rate at the metering intersection (Figure 8.12:).

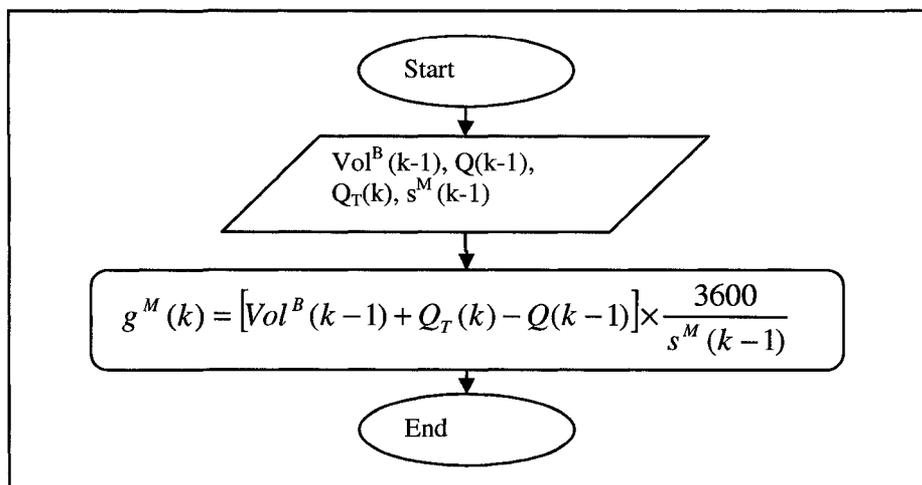


Figure 8.12: Queue Management Algorithm for the Model Based Control

8.6.2 Control Performance of the Model Based Algorithm

Roadway configuration, signal timing parameters at intersection B, the approach volume, and the target queue length Q_T , are identical to the previous examples. The performance of the control algorithm is shown in Figure 8.13 and Table 8.4.

In terms of accuracy and responsiveness, the Model based control is performing quite well. It maintains the target queue length with very little error and responds to target queue Q_T changes within one cycle.

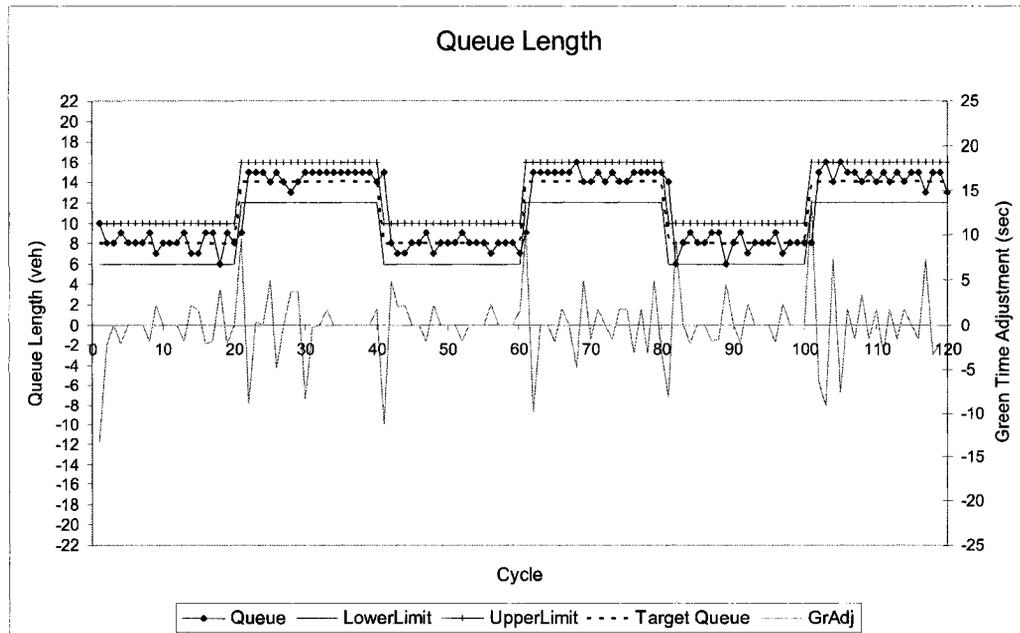


Figure 8.13: Model Based Queue Control Algorithm Output

Table 8.4: Control Performance for the Model Based Algorithm

Accuracy # of points outside $Q_T \pm \tau$	Responsiveness Reaction time delay	RMSE	$\sigma(\text{GrAdj})$
5	1,1,1,1,1	1.523	3.873

The excellent performance of the model based control in the simulation environment is not surprising. The control model is relatively simple thus the mathematical relationships are well defined, and the input data are accurate and reliable. Therefore, predicting the required green time at the metering intersection in a simulation environment can be

carried out without major difficulties. Implementation in a real-world environment is however more challenging (Chapter 10).

8.7 Partial Queue Jump with Fuzzy Logic based Control

The geometric-set up for Fuzzy Control is very similar to the set-up for the Model based control. The only difference is in the type of data collected. Fuzzy Logic control requires only check-in and check-out detector counts at the beginning and end of the detection zone.

8.7.1 Algorithm Description

The general control concept is based on a feedback control loop. The queue in the restricted zone is measured once a cycle, the error and the change of error are calculated, and the green time from the previous cycle is adjusted Figure 8.14.

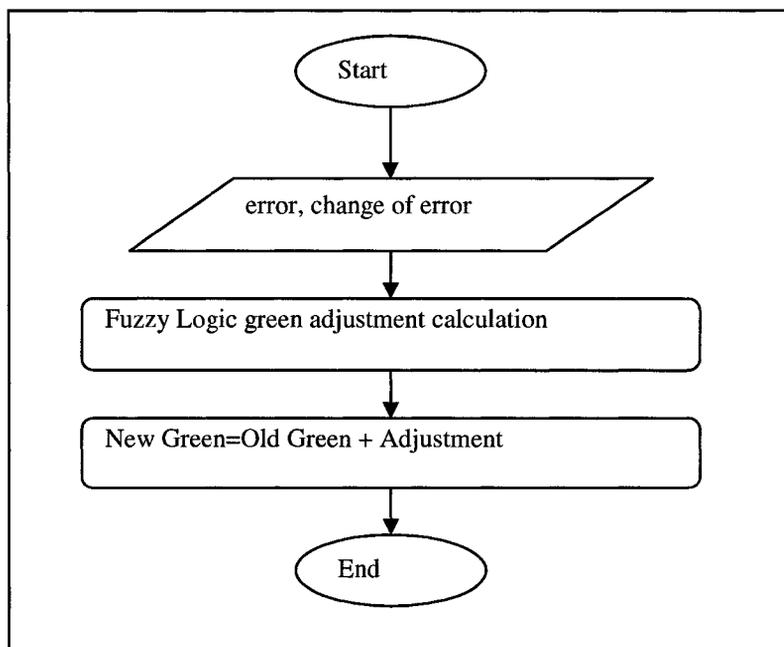


Figure 8.14: Queue Management Algorithm for the Fuzzy Logic Based Control

The development of a general Fuzzy Logic (FL) based queue control concept is provided in Chapter 6. In this section, the FL control algorithm from Chapter 6 is adapted to the calculation of the green adjustment. The algorithm of the Fuzzy Logic based green adjustment calculation is presented in Figure 8.15.

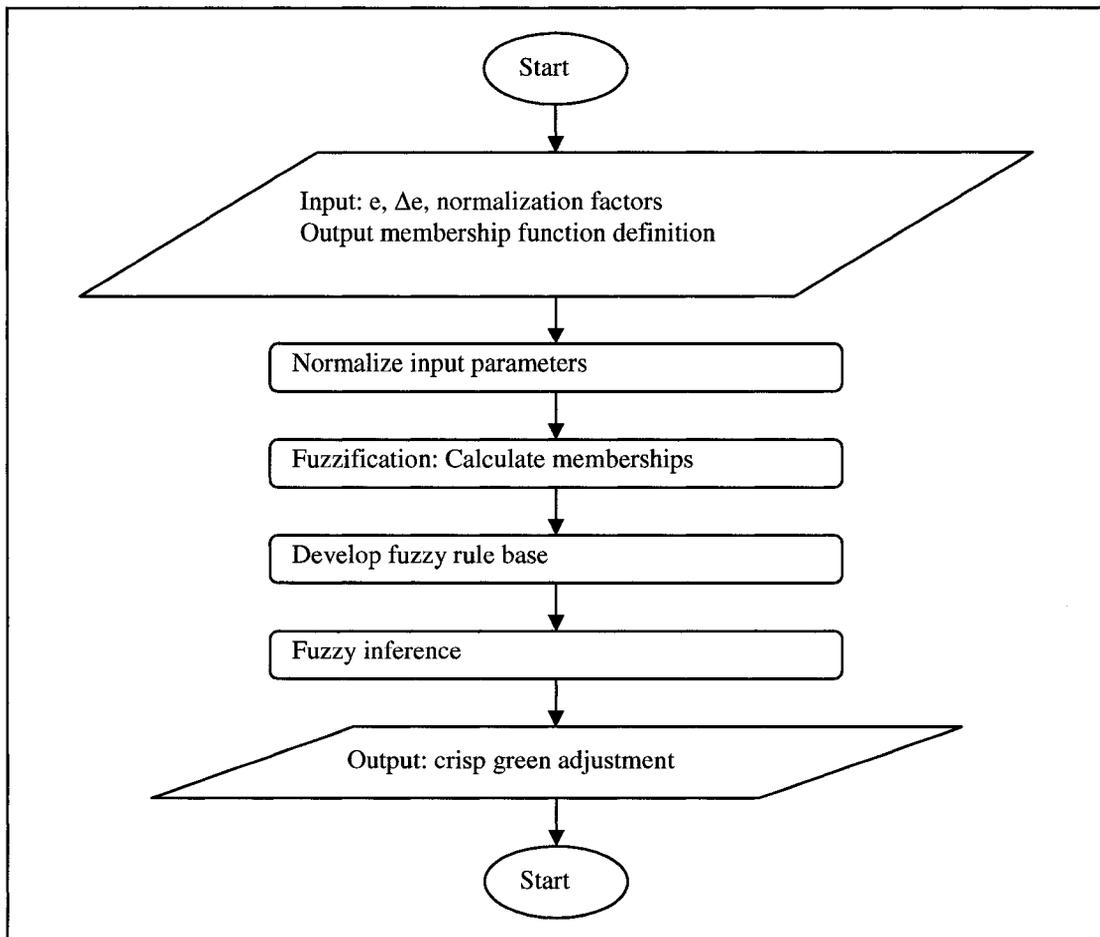


Figure 8.15: Green Time Adjustment Calculation at the Metering Signal Based on Fuzzy Logic

Input data

The algorithm uses two input variables, e is representing the *error*, the difference between the target queue length Q_T and the measured queue length Q_M , and Δe representing the *change of error* between two sampling periods. The interpretation of positive and negative values of the error and change of error is described in Figure 8.16.

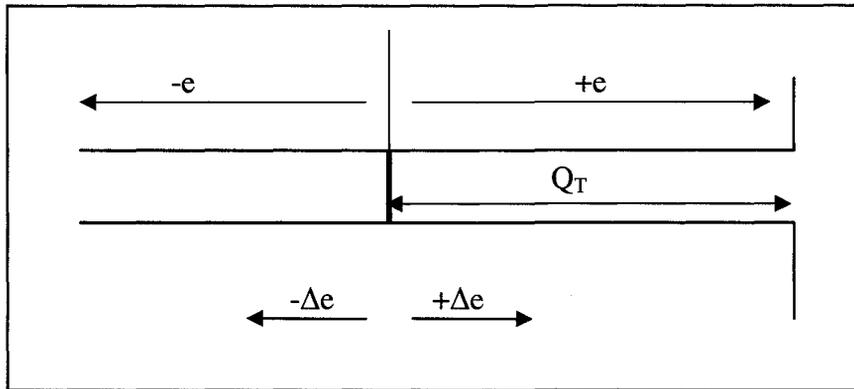


Figure 8.16: The Meaning of Positive and Negative Error and Change of Error

Normalize input variables

Both input variables are normalized as described in Section 6.4.2. The objective of normalization is to simplify the process of fuzzification by using a generic programming code. The normalization coefficient is calculated by Equation 8.10.

$$n_e = n_{\Delta e} = \frac{1}{\max |Q_T - Q_M|} \quad 8.10$$

Fuzzification

Fuzzification is carried out by using the membership functions defined in Table 8.5.

Parameters sf_e and $sf_{\Delta e}$ are the scaling factors defined in Section 6.5.2.1 and they are used for tuning the membership functions.

Rule-base development

The rule base development is presented descriptively in Table 8.7. The condensed description of rules for simulation programming purposes is summarized in Table 8.8.

Table 8.5: Normalized and Scaled Membership Function Parameters for the Error e

Linguistic Variable	e (normalized)			
	a	b	c	d
NL (-2)	$-99/sf_e$	$-95/sf_e$	$-1/sf_e$	$-0.5/sf_e$
NS (-1)	$-1/sf_e$	$-0.5/sf_e$	$0/sf_e$	
Z (0)	$-0.5/sf_e$	$0/sf_e$	$0.5/sf_e$	
PS (1)	$0/sf_e$	$0.5/sf_e$	$1/sf_e$	
PL (2)	$0.5/sf_e$	$1/sf_e$	$95/sf_e$	$99/sf_e$

Table 8.6: Normalized and Scaled Membership Function Parameters for the Change of Error Δe

Linguistic Variable	Δe (normalized)			
	a	b	c	d
NL (-2)	$-99/sf_{\Delta e}$	$-95/sf_{\Delta e}$	$-1/sf_{\Delta e}$	$-0.5/sf_{\Delta e}$
NS (-1)	$-1/sf_{\Delta e}$	$-0.5/sf_{\Delta e}$	$0/sf_{\Delta e}$	
Z (0)	$-0.5/sf_{\Delta e}$	$0/sf_{\Delta e}$	$0.5/sf_{\Delta e}$	
PS (1)	$0/sf_{\Delta e}$	$0.5/sf_{\Delta e}$	$1/sf_{\Delta e}$	
PL (2)	$0.5/sf_{\Delta e}$	$1/sf_{\Delta e}$	$95/sf_{\Delta e}$	$99/sf_{\Delta e}$

Fuzzy inference

Finally, the outputs of individual rules are calculated, the outputs of all rules are combined and defuzzified by using the Takagi-Sugeno type inference system.

The output of the calculation is the green adjustment to green phase from the previous cycle at the metering intersection. The new green time calculated as the sum of the previous green time and the adjustment is then used at signal M.

The simulation model framework is depicted in Figure 8.17.

Table 8.7: Rule Base for Queue Management with Pre-Signal

Queue behaviour	Rule	Control action
I. Queue is less than target and decreasing or more than target and increasing: intensive decrease/increase required	1	IF e is NL AND Δe is NL THEN NL
	7	IF e is NS AND Δe is NS THEN NL
	19	IF e is PS AND Δe is PS THEN PL
	25	IF e is PL AND Δe is PL THEN PL
II. Error exists but it is not changing: decrease/increase with corresponding intensity required	3	IF e is NL AND Δe is Z THEN NL
	8	IF e is NS AND Δe is Z THEN NS
	18	IF e is PS AND Δe is Z THEN PS
	23	IF e is PL AND Δe is Z THEN PL
III. Error is large and decreasing slowly: intensive decrease/increase required	4	IF e is NL AND Δe is PS THEN NL
	22	IF e is PL AND Δe is NS THEN PL
IV. Error is small and diminishing: no action required	9	IF e is NS AND Δe is PS THEN Z
	17	IF e is PS AND Δe is NS THEN Z
V. Error is getting worse: intensive decrease/increase required	2	IF e is NL AND Δe is NS THEN NL
	6	IF e is NS AND Δe is NL THEN NL
	20	IF e is PS AND Δe is PL THEN PL
	24	IF e is PL AND Δe is PS THEN PL
VI. Error is large but decreasing fast: support the change with moderate decrease/increase	5	IF e is NL AND Δe is PL THEN NS
	21	IF e is PL AND Δe is NL THEN PS
VII. Small error and decreasing fast: reverse action to prevent/minimize overshoot	10	IF e is NS AND Δe is PL THEN PS
	16	IF e is PS AND Δe is NL THEN NS
VIII. Error is zero: no action required	13	IF e is Z AND Δe is Z THEN Z
IX. Error is zero but it is changing fast: moderate decrease/increase required to reverse change	11	IF e is Z AND Δe is NL THEN NS
	15	IF e is Z AND Δe is PL THEN PS
X. Error is zero but changing slowly: wait, no action required	12	IF e is Z AND Δe is NS THEN Z
	14	IF e is Z AND Δe is PS THEN Z

Table 8.8: Rule-Base Described by Linguistic Numeric Variables

Green adjustment			Change of error, Δe				
			Queue increasing		Queue decreasing		
			-2	-1	0	1	2
Error, e	Longer than target	-2	-2	-2	-2	-2	-1
		-1	-2	-2	-1	0	1
		0	-1	0	0	0	1
	Shorter than target	1	-1	0	1	2	2
		2	1	2	2	2	2

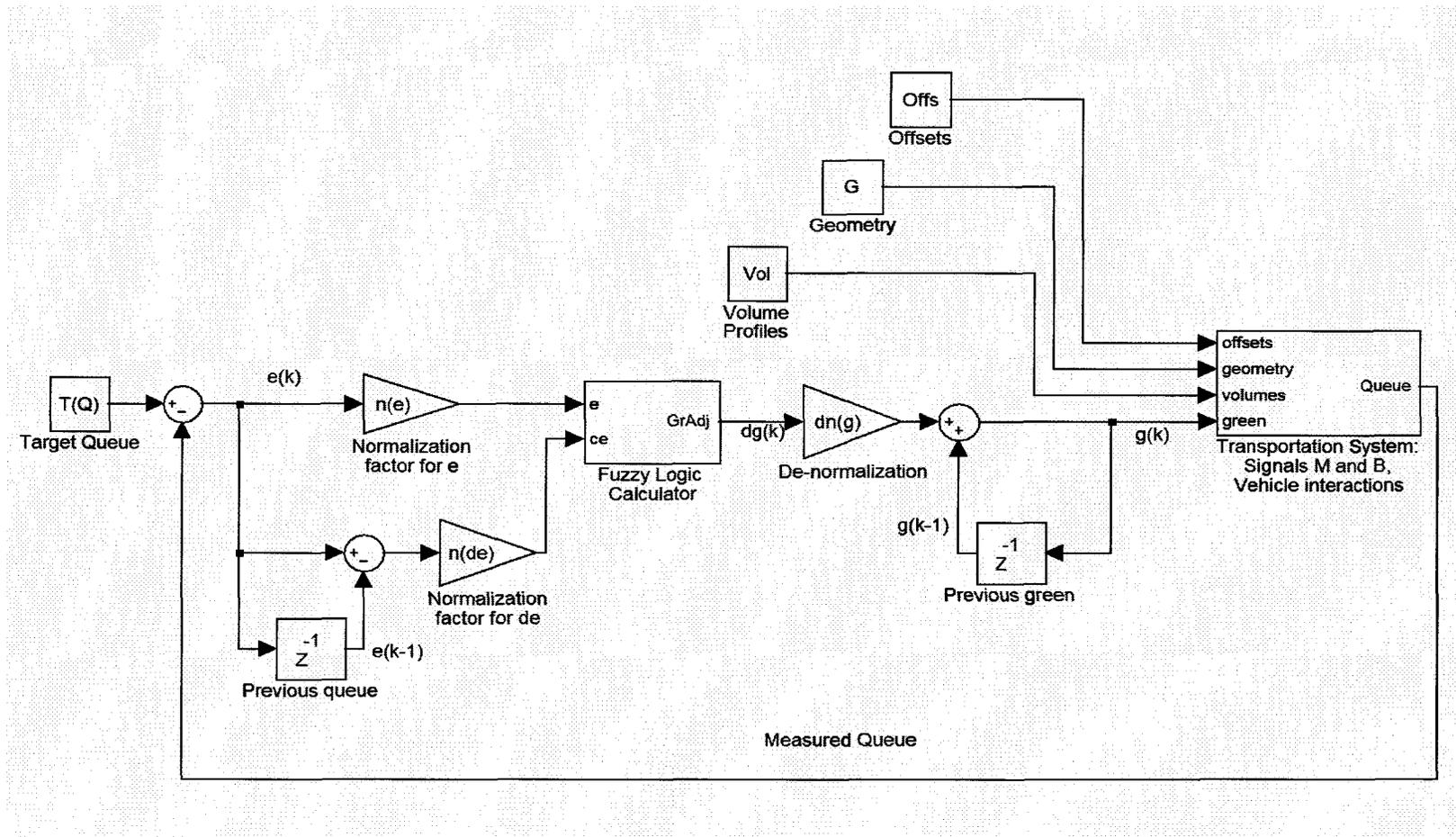


Figure 8.17: Fuzzy Logic Simulation Model Framework

8.7.2 Tuning the Fuzzy Inference System

Tuning a fuzzy system could potentially become very complex exercise due to the large number of tuneable parameters. However, the knowledge about system operation and of the desired output significantly reduces the complexity of tuning. Furthermore, by using scaling factors for the input and output membership function parameters the tuning process is further simplified. Nevertheless, tuning the Fuzzy Inference System is still based on trial and error thus it is time consuming and as much an art as science.

The role of FIS tuning in the queue management design process is summarized in Figure 8.18. The tuning process is an integral part of system design. Changing some of the system parameters, for example the number of lanes or the length of the restricted zone (space available for queuing between the bottleneck and metering intersections) may require re-tuning. However, tuning the FIS is an off-line activity. Once the FIS is tuned during the design stage of the controller algorithm development, it does not require re-tuning as long as the geometry of the roadway infrastructure and the control objectives remain unchanged.

The diagram depicted on Figure 8.19 provides an overview of the data flow during the simulation and tuning process with references to appropriate equations discussed earlier in this thesis.

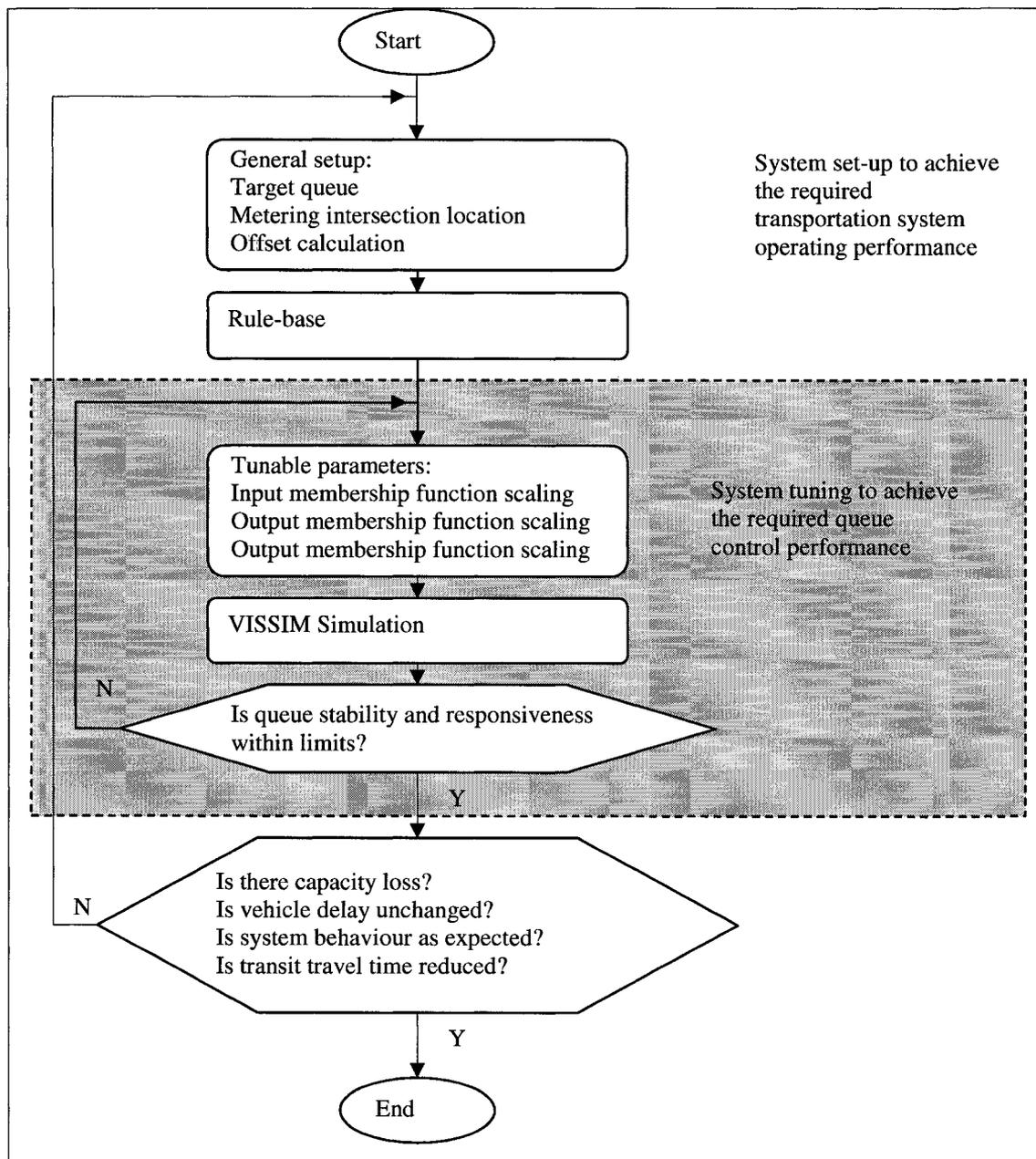


Figure 8.18: The Role of Tuning in the Queue Management Design Process

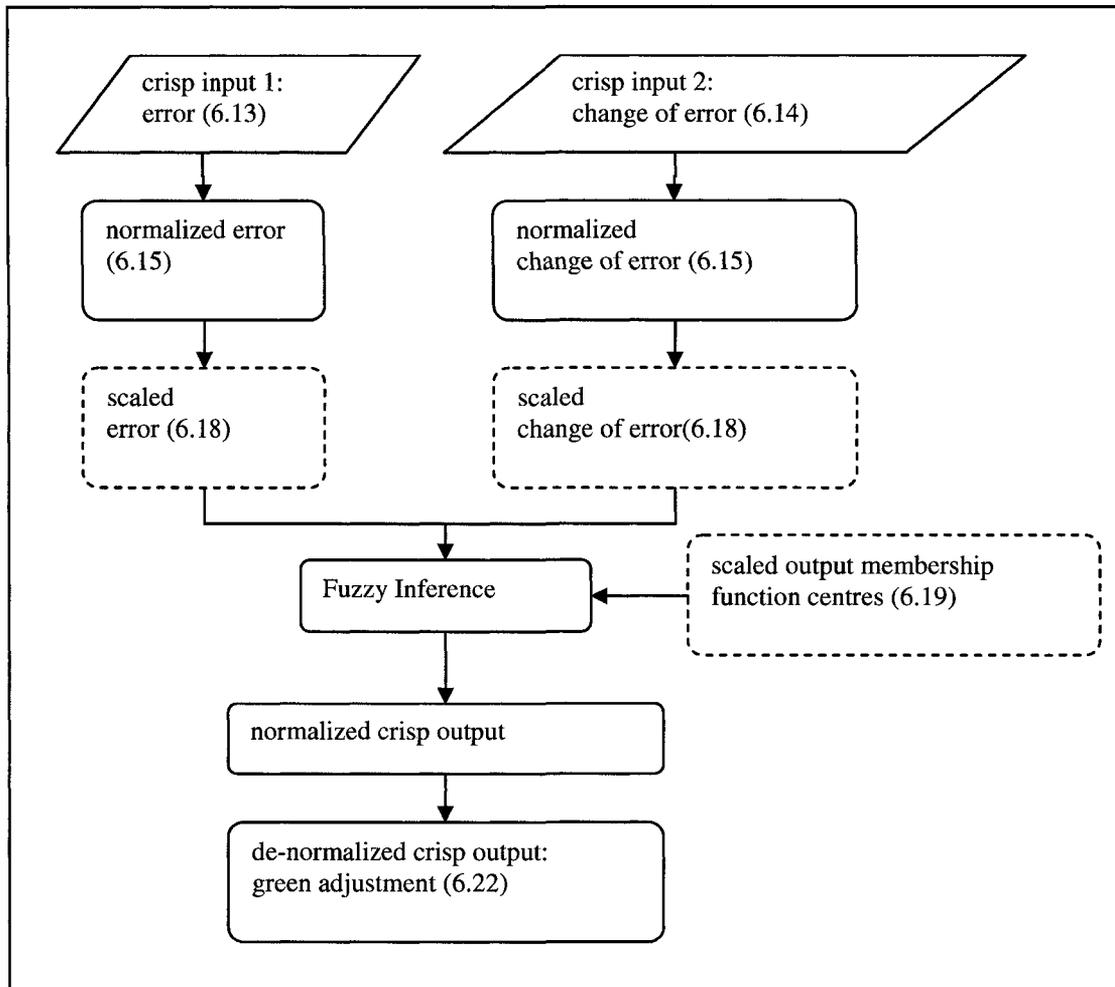


Figure 8.19: Data Flow during the Tuning Process

Similarly to the “ParkView” type control tuning has primary and secondary objectives.

Tuning has been carried out by trial and error. Through experimentation, the following three step procedure has been developed and adopted:

- set the maximum green adjustment (dn_i ; Section 6.5.2.2) to achieve the desired responsiveness;
- scale the output membership function centres to achieve accuracy (minimize oscillation); and

- adjust the scaling factors sf_1 for the error and sf_2 for the change of error to minimize the variability of the control output, green adjustment.

The above procedure will find one of the local optimums and not necessarily the global optimum. However, as discussed previously (Section 8.4.2), for the control system to achieve its operational objectives, the queue management subsystem does not necessarily have to operate at its global optimum.

8.7.3 Control Performance of the Fuzzy Logic Based Algorithm

Roadway configuration, signal timing parameters at intersection B, the approach volume, and the target queue length Q_T , are identical to the previous examples. The performance of the control algorithm is shown in Figure 8.20 and the scaling parameters and performance measures in Table 8.9.

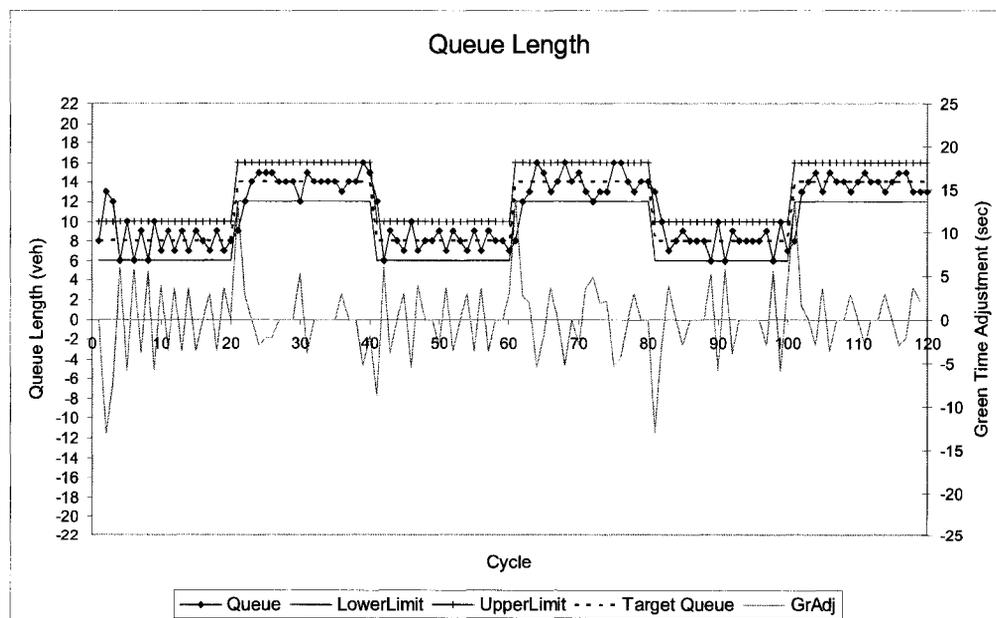


Figure 8.20: Fuzzy Logic Based Queue Control Algorithm Output

In terms of accuracy and responsiveness, the Fuzzy Logic based control is performing acceptably. After the initial learning period it maintains the target queue length within the allowable range and responds to target queue Q_T changes within one cycle.

Table 8.9: Tuned Parameters and Control Performance for the Fuzzy Logic Based Control

Max Green Adj.	Output Scaling factor	Input scaling factors	Accuracy # of points outside $Q_T \pm \tau$	Responsiveness (Cycles)	RMSE	$\sigma(\text{GrAdj})$
dn=16	$\text{sf}_R=0.5$	$\text{sf}_1=1$ $\text{sf}_2=0.4$	7	1,1,1,1,1	3.984	1.526

8.8 Comparison of Control Algorithms for Queue Management with Pre-Signal

The objective is to compare the performance of the Base Case Scenario (do nothing) with the queue management methods by using the four pre-signal control algorithms. The comparison is carried out from the following aspects:

- single mode (traditional traffic engineering) perspective; and
- bimodal transportation system perspective.

The other two evaluation perspectives, transit system operations and transportation policy/planning perspectives (Section 5.4) are demonstrated in the real-world example (Section 8.9.4).

8.8.1 Geometric Set-up and Input Data

The geometric set-ups for all control alternatives are identical in terms of the lane arrangements, and the only minor differences are in terms of detector locations and the type of data that is collected. The set-ups and data collection points are presented in the following figures:

- Base Case Scenario (BCS) - Figure 8.1;
- Queue Management without Signal Control - Figure 8.2;
- “Park View” control - Figure 8.6; and
- Model and Fuzzy Logic based control - Figure 8.11.

As discussed in Chapter 5, to evaluate operations control performance in oversaturated conditions, it is not enough to use average arrival flows during the observation period. To model delay correctly it is necessary to use the arrival (demand) distribution during the observation period. The observation period must be long enough to start and end with arrival rates below capacity. For comparison purposes, a simple arrival profile was used (Table 8.10). To eliminate the impact of the start-up period while the simulated road section is populated, data was not collected during the first 120 seconds. To eliminate the impact of left over vehicles in the system at the end of the simulation, arrivals have been reduced to zero for the last 600 seconds of simulation and data was collected for all vehicles until they left the system (Table 8.10).

Table 8.10: Arrival Volume (Demand) Distribution

Time Period (simulation sec)	EB Vol (veh/hour)
0-120	0
121-1200	350
1200-3600	400
3600-13800	350
13800-14400	0

The target queue length Q_T is 16 vehicles. This number is a function of the location of the pre signal, lane configuration, and signal operating characteristics at the bottleneck intersection and it can be determined by using Equation 8.3.

To evaluate the impact of the queue management function on transit operations, one bus route with a 580 second headway which is slightly less than 10 minutes was added to the system. Normally bus headways are defined in integer minutes, however, 580 sec headway was chosen to ensure that bus arrivals do not occur at the same time within the signal cycle.

The basic performance measure used for comparison is the travel time collected separately for cars and buses. Travel times were collected as vehicles entered the network at the eastbound link and crossed the stop line at the bottleneck intersection. Every scenario was simulated for a period of four hours. Since the VISSIM microscopic simulation is based on vehicle arrivals distributed randomly at the external links (except bus arrivals that are following the pre-set schedule or headway), data collected for the four hour period was averaged and used as one sample travel time measurement. This

sampling was then repeated ten times for each control alternative, every time with a different random number generation seed. The sample means and standard deviations were calculated and used in the comparisons of different alternatives. All alternatives were compared to the Base Case Scenario (control without any queue management).

8.8.2 Evaluation Results

As suggested in Section 8.2.2, the control objectives are two-fold, to maintain existing capacity and to minimize transit delay.

- **Single mode (traditional traffic engineering) perspective**

Vehicle capacity is normally expressed as the maximum number of vehicles that can cross a predetermined point or line during an observation period. The capacity was measured for eastbound vehicles at the bottleneck intersection. The cross section of the roadway approach to the bottleneck intersection and signal timing at the bottleneck intersection did not change and none of the control algorithms starved the bottleneck intersection, thus there was no vehicle capacity change.

Total travel time, including cars and buses, remained very similar to the Base Case Scenario regardless of the control techniques (Table 8.11). The minor variations in travel times are due to randomness but the differences are insignificant based on a statistical comparison of the means (Table 8.12).

When the travel times for all vehicles are disaggregated to car and bus travel times, the advantage of the queue management approach is evident. As expected, transit travel times

decreased substantially with queue management Table 8.12). With the simplest alternative, QMnoPS (queue management without pre-signal) transit travel times dropped to less than half of the BCS and for the other four alternatives to approximately one third of the baseline scenario (Table 8.11).

Also as anticipated, car travel times increased slightly compared to the BCS, but that increase is not significant at the 99% confidence level. A slight increase in car travel times was anticipated since every time there is a change in the sequence of buses and cars (bus jumps the queue), transit vehicles gain a lot in terms of travel time reduction at the expense of an additional small travel time to cars. In terms of total vehicle (cars +buses) travel time, we know from Section 4.3.1.1 that changing the sequence of vehicles has some impact. Simulation results also indicate an increase but based on the 10 simulation runs, the change is not statistically significant.

Table 8.11: Travel Times Generated by Baseline Scenario and Alternative Set-Ups

		Mean Vehicle Travel Times (sec/veh)					
		BCS	QMnoSC	“ParkView”	“On-Off”	Model	FL
Car	Mean	412	439.2	434	423	422	419
	σ	44.3	59	76	57	41	52
Bus	Mean	424	204	152	139	142	140
	σ	43	6.7	5	2	0.4	3
Car+Bus (Total)	Mean	412	434.9	429	418	417	414
	σ	44	58	75	56	40	51.5

(More details are provided in Appendix J)

Table 8.12: The Impact of Queue Management on Mean Vehicle Travel Time

	Change in Mean Vehicle Travel Time Compared to Base Case Scenario (sec/veh)				
	QMnoSC	“ParkView”	“On-Off”	Model	FL
Car	27	22	11	10	7
Bus	-221**	-272**	-285**	-283**	-284**
Car+Bus (Total)	23	17	5	5	2

** Significant at the 99% confidence level

To visualise the above information, Figure 8.21 was created. The figure clearly demonstrates that car and total travel times remained the same, however, bus travel times decreased substantially with all five queue management alternatives.

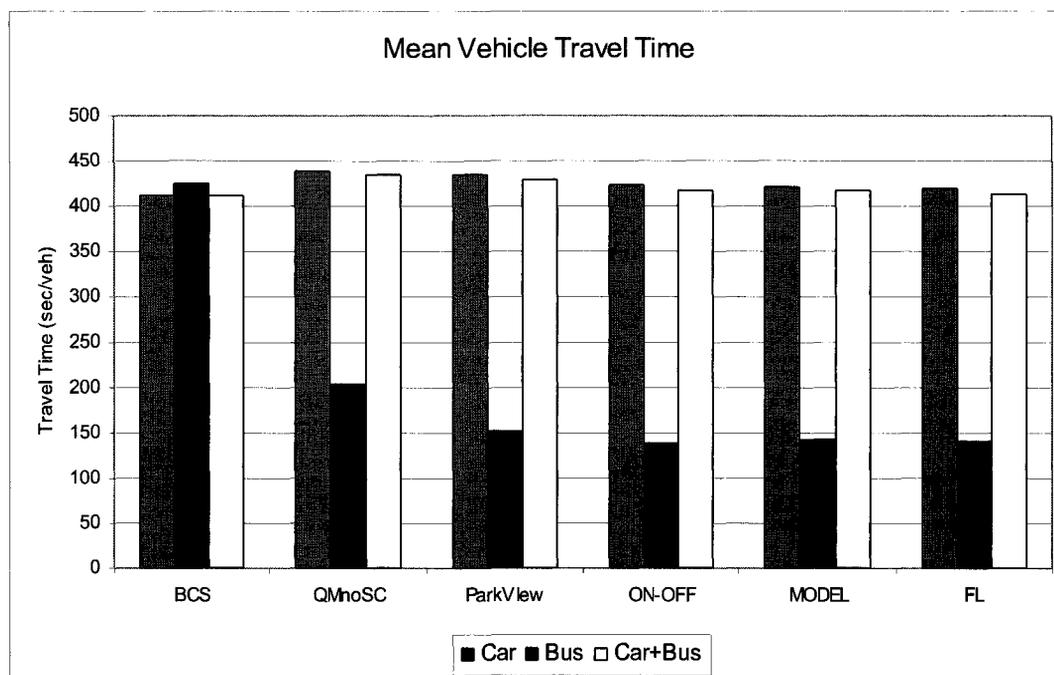


Figure 8.21: Visual Presentation of Travel Time Comparisons of Alternative Controls

Bimodal transportation system perspective

Total person delay was calculated based on vehicle delays and the corresponding vehicle occupancies. For vehicle occupancy values 1.2 and 40 persons per vehicle were assumed for the car and transit modes respectively.

As Table 8.13 demonstrates, the decrease in total person delay is substantial, it is approximately 100 sec/person or 25 % reduction for the alternatives with pre-signals. The reduction with the alternative without a pre-signal (QMnoPS) is slightly less but still quite important. The same information is depicted graphically for better visualisation Figure 8.22.

Table 8.13: Multi-Modal Transportation System Perspective – Impact on Person Travel Time

		Mean Person Travel Times (sec/per)					
		BCS	QMnoSC	“ParkView”	“On-Off”	Model	FL
Car+Bus (Total)	Mean	417	350	327	315	316	314
	σ	44	38	47	35	25	32

Table 8.14: The Impact of Queue Management on Mean Vehicle Travel Time

	Change in Mean Person Travel Time Compared to Base Case Scenario (sec/per)				
	QMnoSC	“ParkView”	“On-Off”	Model	FL
Car+Bus (Total)	-67**	-89**	-101**	-101**	-103**

** Significant at the 99% confidence level

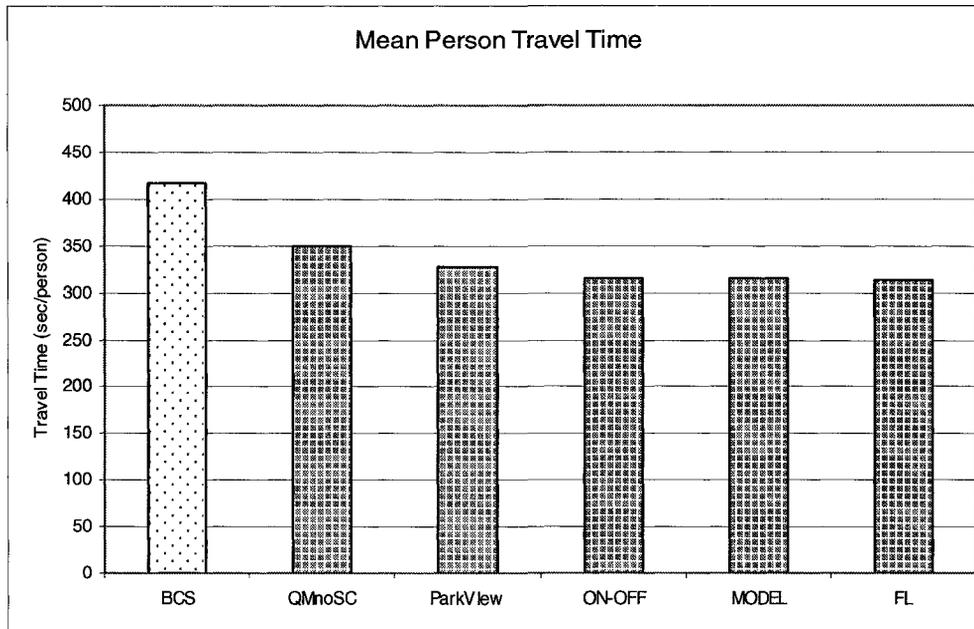


Figure 8.22: Multi-Modal Transportation System Perspective – Impact on Person-Travel Time

8.8.3 Discussion

All five queue management algorithms tested in a simulation environment show improvements in terms of mean transit travel times and mean person travel times compared to the traditional control that has no queue management.

In terms of performance measures, the queue control algorithms that require the installation of a pre-signal (“Park-View” type, “On-Off” type, Model and Fuzzy Logic based control) are somewhat better compared to the queue management approach without pre-signal (QMnoSC) but they are also more costly to implement and maintain.

In real-world, a queue management scheme can function only if motorists observe the rules. Thus, the selection of the queue management approach must consider whether or

not the bus lane designation has to be enforced by design or by legal measures only. If the transit lane cannot be enforced effectively by design (physical measures or signals), the queue management approach without pre-signal will not provide the expected benefits.

The queue management control algorithms were developed and tested in a controlled simulation environment. In this ideal setting, the collection of any input data that is required for controller operation is feasible and relatively reliable. However, in real-world applications, reliable and accurate data collection is much more difficult. The most reliable vehicle detection method is by inductive loops imbedded in the pavement, but even this method has less than 100% accuracy. Furthermore, the loops can get damaged by pavement shift or construction activities, connections can get loose causing intermittent or permanent failures.

From the reliability perspective, the queue management approach that requires the least number of detectors is preferred. Table 8.15 summarizes the detection requirements of each queue management alternative. The Model based queue management requires the most input from the field, has the highest requirement for detectors, thus it is likely the least reliable. It may be possible to use some detectors for more than one function, for example the check-out detector at the bottleneck intersection could be used for measuring volume as well.

Table 8.15: Detection Hardware Requirements

Method	Data Requirement	Data Collection
No Queue Management	No requirement	No collection
QM no PS	No requirement	No collection
“ParkView”	Queue length	1 check-in detector per lane 1 check-out detector per lane
“On-Off”	Queue length	1 check-in detector per lane 1 check-out detector per lane or one occupancy detector per lane
Model based	Queue length Saturation flow rate at M Volume count at B	1 check-in detector per lane 1 check-out detector per lane 1 detector per lane 1 detector per lane
FL	Queue length	1 check-in detector per lane 1 check-out detector per lane

One important aspect of reliability is what happens when some of the loops do not provide accurate data. Total detection failure is a smaller problem because it is easier to identify it. Intermittent detection problems or systematic biased data is a bigger issue. For example, systematic overcounting at a check-in detector results in overestimated queue lengths and a gradual failure of the queue management algorithms.

Vehicle detection by GPS or other Intelligent Transportation System (ITS) technologies are already available. It is quite likely that these technologies could be adapted in the near future to provide the required input data for queue management algorithms.

Nevertheless, input information accuracy and reliability remains to be a major concern because the queue calculation method is very sensitive to systematic errors. Two algorithms, the “On-off” and the Fuzzy Logic type can be relatively easily modified to

The abrupt termination of the arterial roadway at the vicinity of to the residential neighbourhood is quite typical in Ottawa, and it reflects the society's resistance to new roadways in built up urban areas. A few decades ago, the plan was to extend Carling Avenue as a six lane cross section further east, and widen Bronson Avenue to a six lane cross section North and South. However, due to the insurmountable public resistance, the plan has been abandoned.

The intersection of Carling Avenue and Bronson Avenue represent one of the bottlenecks in the area. In the Northbound direction, the intersection is one of many bottlenecks along Bronson Avenue, but in the eastbound direction, the intersection is the main bottleneck. Nevertheless, there are no short or long term plans to increase the vehicular capacity of the intersection. The intersection has been recently modified (2005) with the objective to improve the pedestrian environment, the safety of the northbound left turning movements, and the eastbound transit movement. Increasing vehicular capacity was not an objective.

The implemented transit improvement can be described as simplified bimodal queue management technique, and in this evaluation, it is referred to as Scenario A. Scenario B represents further improvements, and the "Before" Scenario pertains to the "traditional" operation without any queue management measures.

Figure 8.24 depicts the eastbound approach to the intersection for the "Before" Scenario. The critical transit route runs eastbound along Carling Avenue in the mixed flow

environment, and after servicing the bus stop at Cambridge Street, the bus leaves the curb lane and turns left at Bronson from the middle lane. Vehicular queues regularly build up in the eastbound direction during both peak periods. Since the eastbound direction is leaving the bottleneck intersection, there are no eastbound capacity constraints within the observed section.

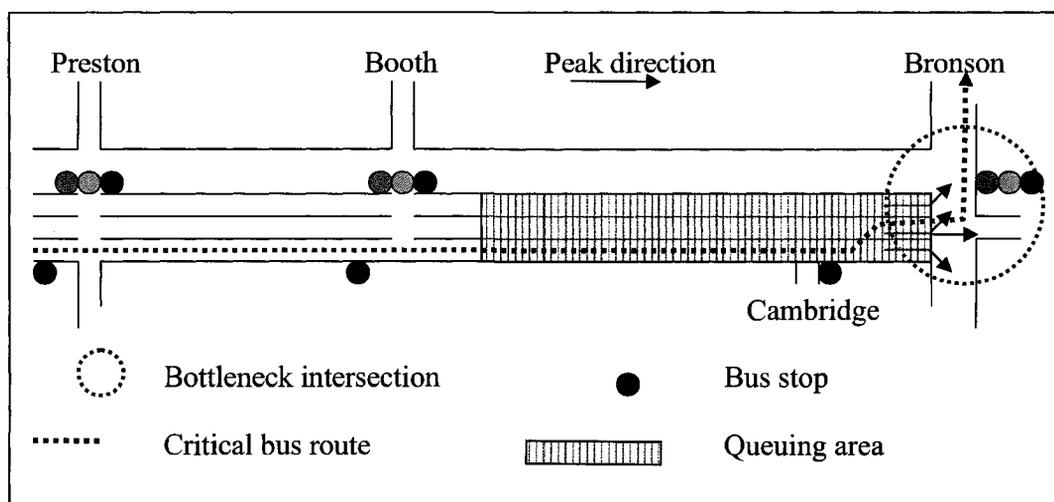


Figure 8.24: Schematic Lane Arrangement for the “Before” Scenario in the Critical Eastbound Direction

The operations analysis of the corridor is relatively simple. The seven step analysis process proposed in Section 7.3.2.2 is followed. The first four steps of the seven step analysis process described in Figure 7.8 are trivial. The location of the bottleneck in the observed eastbound direction is obvious, and there is no other secondary bottleneck all the way to the intersection of Carling Avenue and Preston Street. The space between these two major intersections is normally sufficient to store the queues.

Step 1: Volume balancing

Traffic counts are available for all three signalized intersections within the observed section. Since there is one prominent bottleneck intersection in the eastbound direction (i.e. Carling Avenue and Bronson Avenue) traffic volume counts taken at this intersection were used as the basis for volume balancing.

Step 2 and Step 3: Determine signal operation and fine tune signal and operations parameters

Signal operation at all three intersections within the observed section is relatively simple. Timing plans and stage sequencing is summarized in Appendix K.

Step 4: Identification of capacity bottleneck in the eastbound direction

The location of the capacity bottleneck is known from experience. It is also confirmed analytically (Appendix K).

Step 5: Simulation by using arrival distribution

Although traffic counts collected by the City of Ottawa at the intersection of Carling Avenue and Bronson Avenue pertain to 15 minute intervals, they do not reflect the through arrival distribution because they were collected at the stop bar as vehicles entered the intersection. The arrival profile can be collected by counting vehicles as they leave the upstream intersection which is not impacted by the queues from the bottleneck intersection (e.g. the intersection of Carling Avenue and Preston Street).

Step 6: Fine tuning the impact of the capacity bottleneck

The capacity of the bottleneck intersection is impacted by the duration of the green interval, the cycle length, and the conflicting pedestrian movements. Only one of these three factors, the conflicting pedestrian movement, is variable at this intersection.

Although one pedestrian arrival rate has been used for the duration of the simulation due to the lack of other data, the pedestrian arrivals in VISSIM are randomly distributed thus the number of pedestrian crossings change from cycle to cycle.

In VISSIM, the pedestrian/vehicle conflicts are handled by defining priority rules which prevent vehicles and pedestrians to occupy the same space at the same time. The priority rules were set up to allow pedestrians to cross the road during the green and flashing don't walk pedestrian indications while vehicles waited. Thus the actual vehicle departure (saturation flow rate) was impacted by the pedestrian activity.

In oversaturated conditions, when the arrival rate is larger than the departure rate, queues build up very rapidly. Thus, the arrival volume and arrival distribution on the eastbound approach to the bottleneck intersection were adjusted by trial and error through simulation to prevent queue overflows outside of the simulated link (i.e. assumed distribution was used, Section 7.3.2.2).

Step 7: Data collection for evaluation

The proposed alternatives impact only travel times of vehicles approaching the intersection from the eastbound direction. The origin and destination points for car and bus travel time data collection are defined in Figure 8.30.

Two alternative scenarios are analysed, one without a pre-signal (partial queue jump, and the other with a pre-signal.

8.9.2 Geometric Set-Up and Operation of Partial Queue Jump

without Pre-Signal: Scenario A

The bimodal queue management approach with both scenarios is based on the following logic: If intersections represent the capacity bottlenecks on an arterial roadway, then the links between intersections must have some underutilized capacity. If the difference between the capacity of the link and the bottleneck intersection is sufficiently large, it may be possible to remove a lane from the link and still have sufficient link capacity to feed the intersection.

The capacity of the eastbound approach can be calculated by using Equation 8.6 and it is never more than 1350 veh/hr. (The effective green time at the eastbound approach of the intersection is typically less than 30 seconds, the saturation flow is also not higher than 1800 veh/hr due to the conflicting pedestrian movements.)

$$\frac{30\text{sec}}{120\text{sec}} \times 1800\text{veh/hr} \times 3\text{lanes} = 1350\text{veh/hr}$$

The capacity of one lane at a midblock section is well over 1200 vehicle/hr, two lanes can provide at least 2400 vehicles but the approach capacity is never more than 1350 veh/hr.

The practical purpose of the three lane cross section is to store vehicles. Therefore, one lane is converted to transit use between Preston Street and Cambridge Street. East from the Cambridge bus stop, all three lanes feed into the intersection.

After serving the Cambridge bus stop, buses join the medium mixed flow lane to make a left turn at Bronson Avenue (Figure 8.25). After leaving the curb lane, buses do not receive any priority and are exposed to the same delay as general traffic.

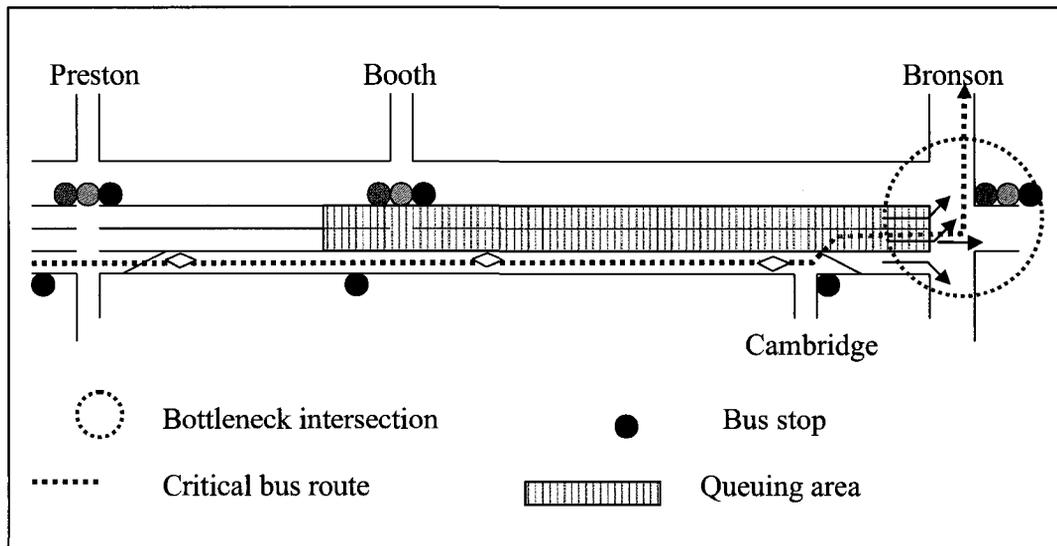


Figure 8.25: Alternative 1 – Partial Queue Jump Without Pre-Signal

Signal timing at the bottleneck intersection is not influenced by the presence of the bimodal queue management facility. Timing is set to satisfy various objectives including vehicle capacity and pedestrian safety criteria, but no special provisions are provided to transit vehicles.

The distance between the end of the bus lane (a point where the three lane cross section for general traffic starts) and the stop bar at the intersection is critical for the effective operation of the facility. This distance can be determined by using the analytical formula presented in Section 8.2.3 to calculate the approach capacity per cycle and by increasing this value by δ (Equation 8.3) to prevent intersection starvation due to variations in the

saturation flow rate. In the general case, the cycle capacity at an approach can vary due to the changes in the length of the green signal and /or due to changes to the saturation flow rate. In case of saturated or oversaturated conditions (which is the case here), normally there is little or no variation of green time because the green intervals are always extended to their maximums. The variation of the cycle capacity at this intersection is caused mainly by the impacts of the pedestrian activity on the saturation flow rate of the right and left turning movements plus the availability of right-turns on red.

For the purpose of implementation, the capacity of the approach per cycle was determined by field measurements. The field measurement is based on vehicle counts that cross the stop bar in each lane per cycle. In this example, the field measurements were used to calibrate the simulation model.

The enforcement of the bus lane designation with a partial queue jump needs to be resolved. The bus lane designation can be enforced by placing a physical barrier (bulb-out) at the end of the bus lane (Figure 8.26).

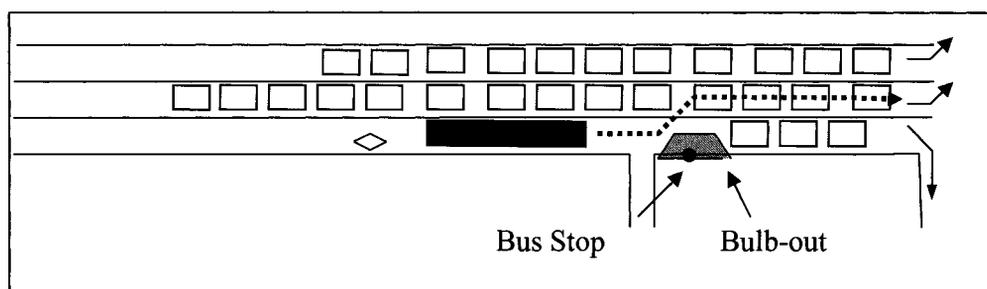


Figure 8.26: Partial Bus Queue Jump Lane with Physical Barrier

Buses turn left at the intersection, therefore their movement is not hindered by the bulb-out. The bulb-out is illustrated in (Figure 8.27).

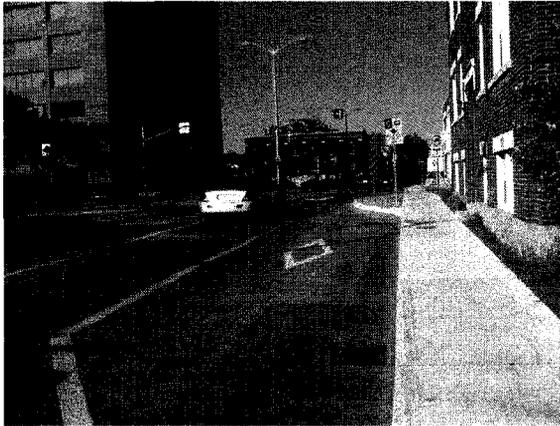


Figure 8.27: Bulb-Out Used to Enforce a Partial Queue Jump Lane

8.9.3 Geometric Set-Up and Operation of Partial Queue Jump with Pre-Signal: Scenario B

Scenario B is based on the previous scenario and a pre-signal is added upstream from the last bus stop. The purpose of the pre-signal is to hold traffic back from the bottleneck intersection and create a 'restricted zone' that allows the bus to enter the bottleneck intersection with less delay. The pre-signal moves the queuing area upstream approximately by the length of the restricted zone (Figure 8.28).

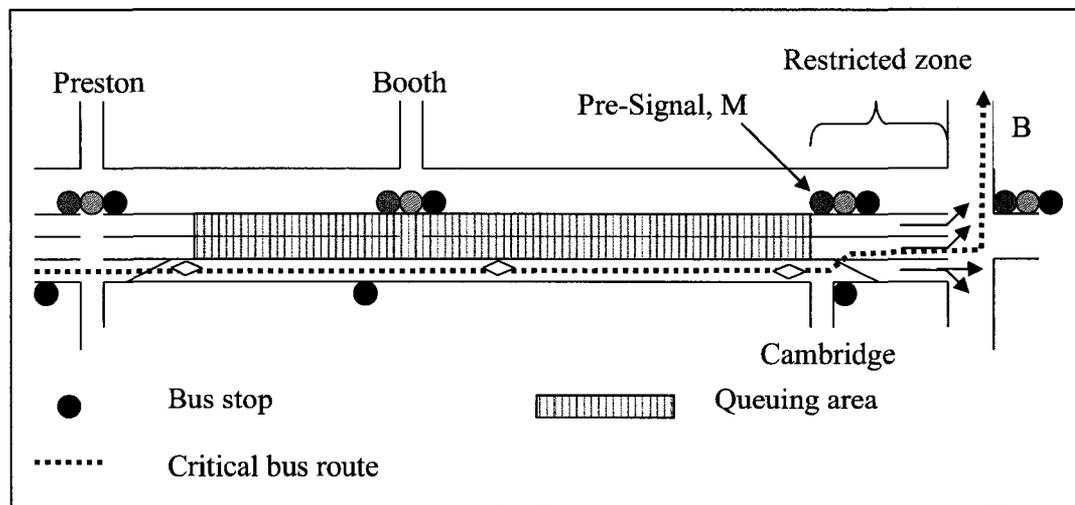


Figure 8.28: Alternative 2 – Partial Queue Jump With Pre-Signal

The signal operation at the bottleneck intersection B is set to satisfy the various control objectives at that intersection, similarly to Scenario A, and it is not influenced by the operation of the pre-signal. The operation of the pre-signal M is coordinated with signal B. Both signals have the same cycle length, and the green starts in the eastbound direction are coordinated through offsets (Section 2.4).

For the control algorithm, the “On-Off” type control (Section 8.5) was selected due to its relative simplicity and least detection requirements. This type of control is feasible at this location, because the pre-signal has only one function, to stop the flow of vehicles when the restricted zone has been filled up. If other functions were required e.g. the control of a pedestrian crossing at the pre-signal, the “On-Off” type control would not be feasible.

The other control algorithms would have to be examined.

Since the control objective of the pre-signal is only to control queue formation, the length of the green interval does not have to be known at the beginning of the signal cycle, it can be terminated when the queue length has reached a pre-set target. The queue length is measured by occupancy detectors d_1 and d_2 , one in each lane (Figure 8.29). When the occupancy detectors are covered for x seconds or longer in both lanes, they would indicate that the target queue length has been reached and the pre-signal can stop the flow of vehicles.

This type of queue measurement is somewhat different from the approach using the check-in/ check-out loops presented earlier in this chapter. In a simulation environment, using the check-in/check out approach is somewhat more straightforward because the unit of measurement corresponds to the unit of measurement of the approach capacity. However, from the perspective of implementation, queue measurement using occupancy detection is more robust. (This is further explored in Chapter 10.)

The pre-signal indications apply only to the general flow lanes. In terms of the bus lane, there are at least two alternative approaches. The first approach is to leave the bus lane uncontrolled. The other approach is to control the bus lane with bus actuated transit signal display. Although the second alternative is technically more complex, it practically eliminates potential bus lane violation by increasing the level of lane use control.

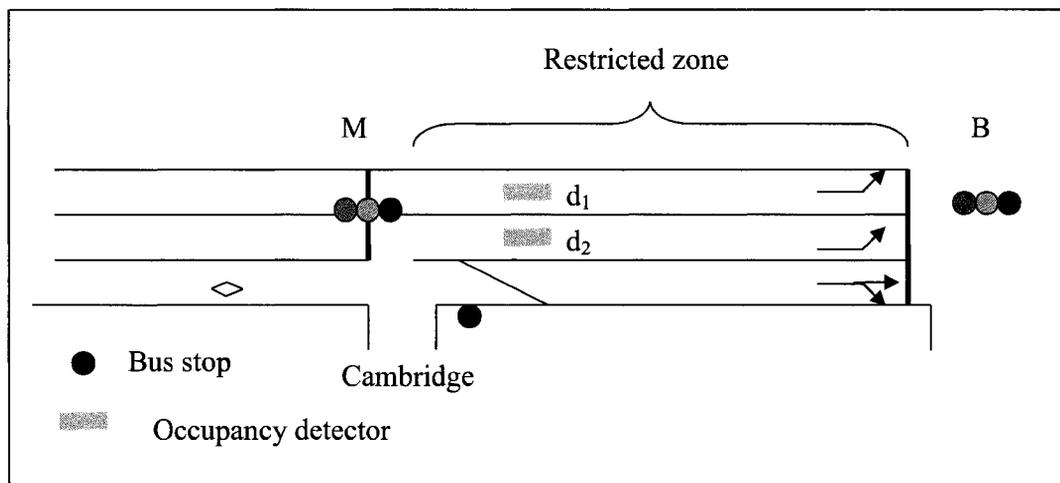


Figure 8.29: Alternative 2 – Pre-Signal Set-Up

8.9.4 Simulation Results

There are a number of ways to present simulation results and the type of presentation is determined by the objectives of the evaluation. The evaluation was carried out from the following perspectives:

- single mode perspective (mean vehicle travel times perspective);
- bimodal transportation system perspective;
- transit system operations perspective; and
- transportation policy perspective.

Three scenarios were compared, the “Before” Scenario, Scenario A – Queue Jump without pre-signal, and Scenario B – Queue Jump with pre-signal. Two hours of operation were simulated, and the simulation was repeated ten times with different random number seeds. The ten runs were treated as ten samples. Travel times were

collected for origin-destination pairs A-B, A-C, and A-D (Figure 8.30). For the A-B pair, travel times for cars and transit are collected separately.

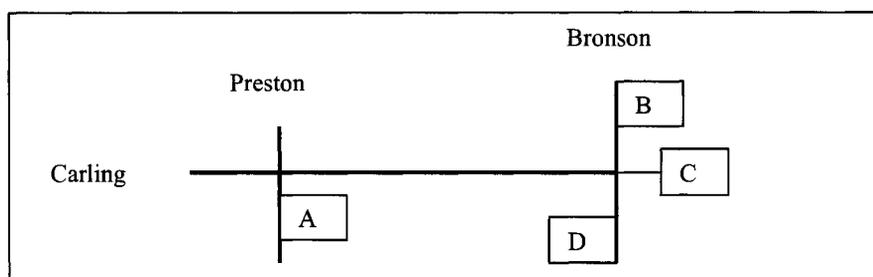


Figure 8.30: Definition of Travel Sections for the Carling Avenue Evaluation

8.9.5 Single Mode Perspective

Travel times were collected separately for cars and buses. Table 8.16 summarizes the mean vehicle travel times and standard deviations for the ten simulation runs.

Table 8.16: Simulation Results: Vehicle Travel Times and its Variability

		Vehicle Travel Times (sec)		
		Before	Scenario A	Scenario B
Car	Mean	420	443	432
	σ	54	39	39
Bus	Mean	401	263	213
	σ	87	5	5
Car+Bus	Mean	420	440	429
	σ	53	38	39

In Table 8.17, the mean travel time differences are presented. For example, by implementing scenario A, the mean car travel time and the total (car+bus) travel time increased slightly, although the change is not statistically significant. However, at the same time, the mean transit travel time decreased significantly by 139 sec. Scenario A transit travel times can be further improved by 50 seconds by adding the pre-signal with

the proposed control logic (Scenario B). Thus the total transit travel time saving is 188 seconds which represents a 47% decrease.

Table 8.17: Comparison of Scenarios – Vehicle Travel Time Perspective

	Vehicle Travel Time Change (sec)		
	“Before” and Scenario A	Scenario A and Scenario B	“Before” and Scenario B
Car	22	-10	12
Bus	-139 (35%)**	-50 (19%)**	-188(47%)**
Car+Bus	20	-11	9

** Significant at the 99% confidence level

A large decrease in the variability of transit travel times is also evident, from 87 to 5 sec.

This is further explored in Section 8.9.7.

It is interesting to examine how travel times change for the three scenarios during the two hour observation period. In the “Before” Scenario, transit vehicles are part of the general traffic flow, thus the transit travel time profile is similar to vehicle travel times for right turning movements, but transit travel times are higher due to the time spent at transit stops (Figure 8.31).

In Scenario A, transit vehicle move freely up to the Cambridge stop (Figure 8.32). Thus the transit travel times are relatively stable. The periodic oscillation is mostly due to the variability of the travel time between the Cambridge stop and the intersection. In many cycles, after leaving Cambridge stop, the bus cannot clear the intersection during the first green signal.

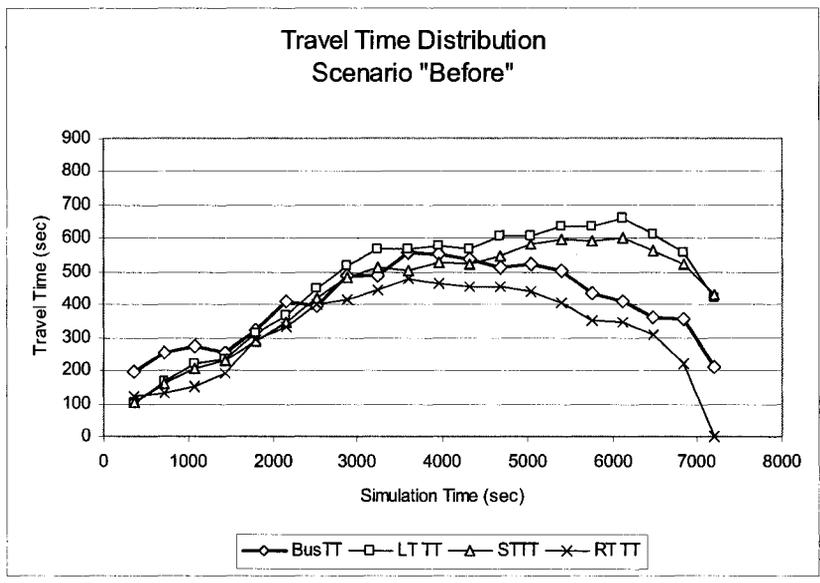


Figure 8.31: Distribution of Travel Times for the “Before” Scenario

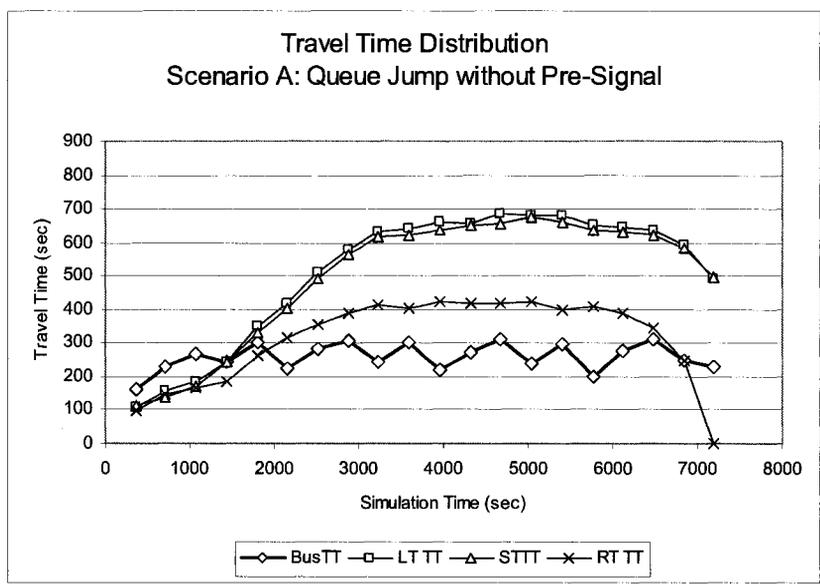


Figure 8.32: Distribution of Travel Times for Scenario A

In Scenario B, transit travel time oscillation is much smaller, because the pre-signal decreases the time when the restricted zone is filled up with vehicles, thus the bus has a much better chance to clear the intersection during the first green interval (Figure 8.33).

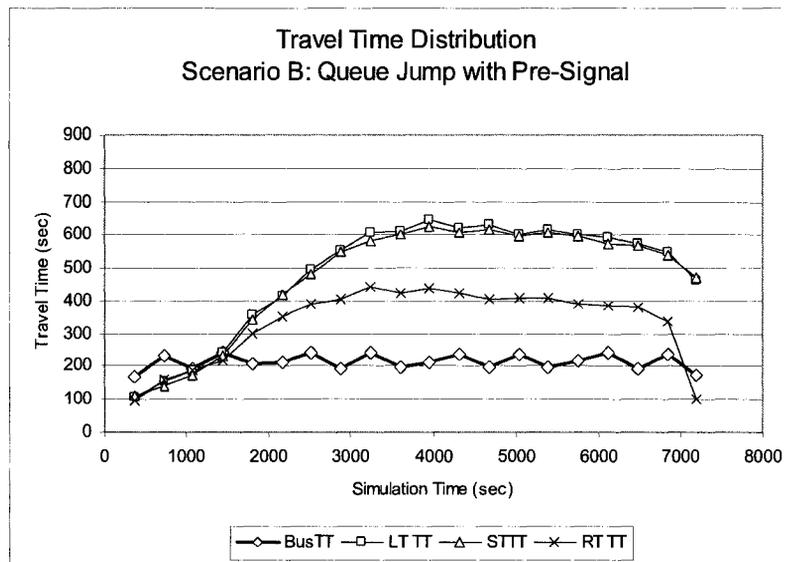


Figure 8.33: Distribution of Travel Times for Scenario B

These results support the conceptual analysis presented earlier in Section 4.3.1.1 that it is possible to decrease transit travel times substantially without significantly impacting car travel times. The conceptual analysis in Section 4.3.1.1 also indicated that by changing the sequence of vehicles (with a bus queue jump), total vehicle delay slightly increases. The increase in total vehicle delay has been observed in the simulation as well, but it is statistically not significant.

8.9.6 Bimodal Transportation System Effectiveness Perspective: Impact on Person Travel Times

The person travel time calculation is based on car and transit travel times and the corresponding vehicle occupancies. No vehicle occupancy data was collected, default values of 40 persons/veh and 1.2 persons/veh were used for the transit and car modes, respectively. Mean person travel times and its variability are presented in Table 8.18 for the three alternative scenarios.

Table 8.18: Simulation Results: Person Travel Times and its Variability

		Person Travel Times (sec)		
		“Before”	Scenario A	Scenario B
Car+Bus (Total)	Mean	414	380	355
	σ	53	25	26

Table 8.19 shows the change in person travel times as the two queue management scenarios are implemented. Scenario A reduces mean person travel time by 34 seconds, and Scenario B further reduces it by 25 seconds. The total reduction is 59 seconds which represents a 14% improvement.

Table 8.19: Comparison of Scenarios: Person Travel Time Perspective

	Person Travel Time Change (sec)		
	“Before” and Scenario A	Scenario A and Scenario B	“Before” and Scenario B
Car+Bus (Total)	-34 (8%)*	-25 (7%)*	-59 (14%)**

* Significance at the 95% confidence level, ** Significance at the 99% confidence level

It can be concluded that when the total vehicle travel times were compared, the queue management increases total travel times. However, in terms of person travel time, the queue management scenarios reduce the total person travel time.

8.9.7 Transit System Operations Perspective

It was proposed earlier (Section 5.4.3) that the mean transit vehicle travel time is a poor performance measure for assessing the impact of queue management on transit operation including schedule design. Let's assume that the 85th percentile travel time is a reasonable basis for schedule calculation. This means that 15 % of buses are not expected to arrive on time at the final stop, and either driver's break time will be curtailed or the next route will start later.

Travel time distributions and the cumulative percentiles for the three scenarios are presented in Figure 8.34, Figure 8.35, and Figure 8.36. The reduction in travel time variability is quite dramatic as the queue management measures become more sophisticated.

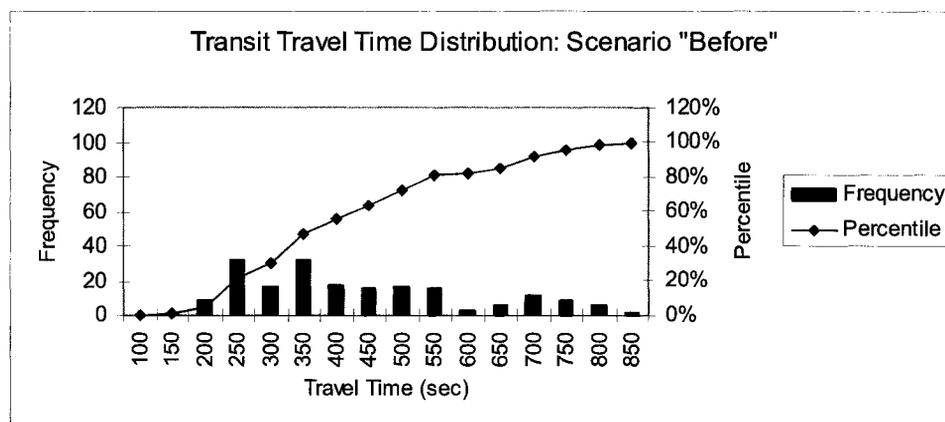


Figure 8.34: Distribution of Transit Travel Times – “Before” Scenario

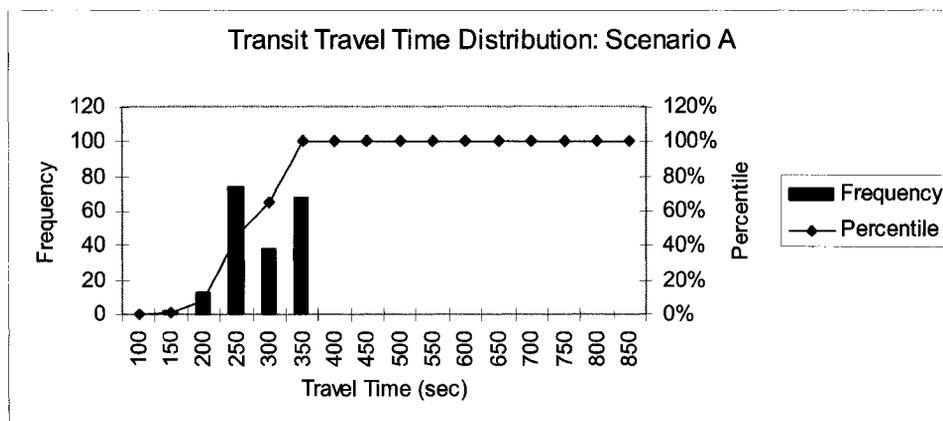


Figure 8.35: Distribution of Transit Travel Times – Scenario A

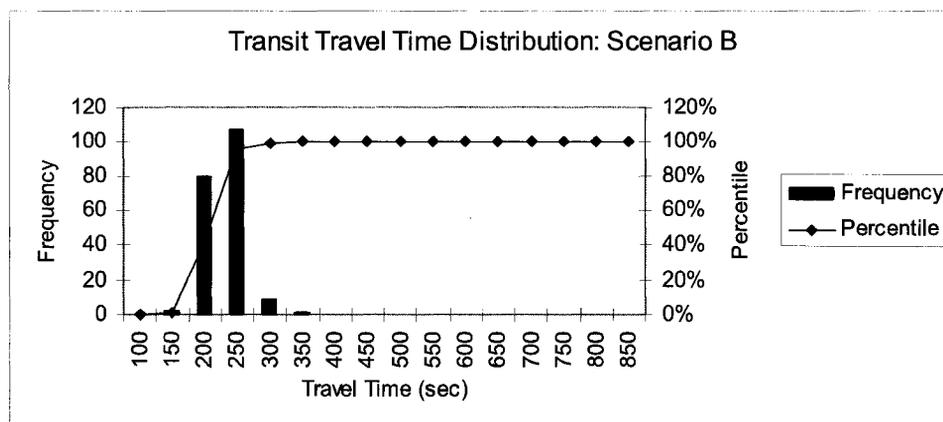


Figure 8.36: Distribution of Transit Travel Times – Scenario B

The 85th percentile and the mean travel times for the three scenarios are summarized in Table 8.21. It is obvious that both the mean and the 85th percentile travel times decrease with Scenario A and Scenario B.

Table 8.20: The Mean and the 85th Percentile Travel Time

	Travel Time (sec)		
	Before	Scenario A	Scenario B
Mean	401	263	213
85 th Percentile	632	310	242

However, the mean travel time underestimates the benefits, because it does not capture the impact of queue management on the distribution of transit travel times. Both queue management scenarios narrow down the distribution of transit travel times dramatically (Figure 8.34, Figure 8.35, and Figure 8.36). The reduction in travel time variability is clearly reflected in the 85th percentile transit travel times.

While the mean travel time saving with Alternative A (compared to “Before”) is 138 seconds, the 85th percentile saving is 322 seconds, which is more than double (Table 8.21). From the transit operations planning perspective, the difference is important. For a transit route with a 5 min service frequency, the 322 second travel time saving means that the number of vehicles required for the service can be reduced by one, or the service frequency can be increased.

Table 8.21: Comparison of Travel Time Savings Expressed by the Mean and the 85th Percentile Travel Time

Travel Time Measure	Travel time savings (sec)		
	“Before” to Scenario A	Scenario A to Scenario B“	Before” to Scenario B
Mean	-138	-50	-188
85 th percentile	-322	-68	-390

It can be concluded that by using mean travel times to assess the impact on transit travel, the benefits of reduced travel time variability are underestimated.

8.9.8 Transportation Policy Perspective

Increasing transit modal share is a typical transportation policy objective in most urban areas in Canada. As discussed in Section 5.3.2, calculating the impact of travel times on transit modal share (TMS) is somewhat problematic. However, ignoring the impact of transit travel time improvements on TMS because there is no simple and accurate estimation method is not a good option either. Therefore, the Transit-Car-Travel-Time ratio (TCTTr) has been proposed as a surrogate performance measure.

Table 8.22 summarises the TCTTr values for the three scenarios. The rationale for using the 85th percentile transit travel time has been already explained. Using the mean for defining car travel times instead of a measure which includes the variability (similarly to transit) needs further discussion. One could argue that, that car travel time variability should be also included in the analysis, similarly to transit passengers, since car travellers are also exposed to the negative impacts of travel time variability. However, there are two reasons why the impact of travel time variability is different for car and transit trips. First, car trips are mostly independent trips. One travels from home to work and the trip is finished with the arrival to the destination. The variability of arrival times has a negative impact on the user, however, it does not impact the following car trip which may be a few hours later. In terms of transit service, the travel time variability between A and B has always an impact on the trip between B to C (Figure 8.37). The transit operator must either provide longer recovery time at B (which impacts capital and operating costs) or the reliability of the service will suffer on all subsequent transit trips.

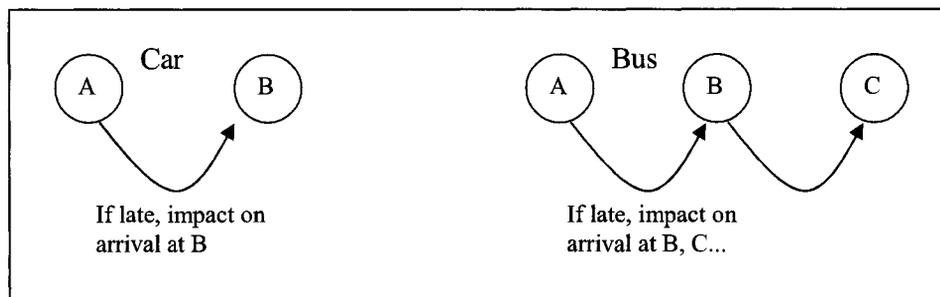


Figure 8.37: The Impact if Travel Time Variability (Lateness) on Car and Transit Trips

The second reasons for using the mean car travel times in evaluations is that from the system perspective those 50 % of travellers that need more than the mean travel time are compensated by the other 50 % that arrive earlier than the mean time (Figure 8.38). If the 85th percentile car travel time is used, 85% of the car trips would be actually shorter. Therefore it seems to be more logical to use the mean car travel times in the TCTTr calculation.

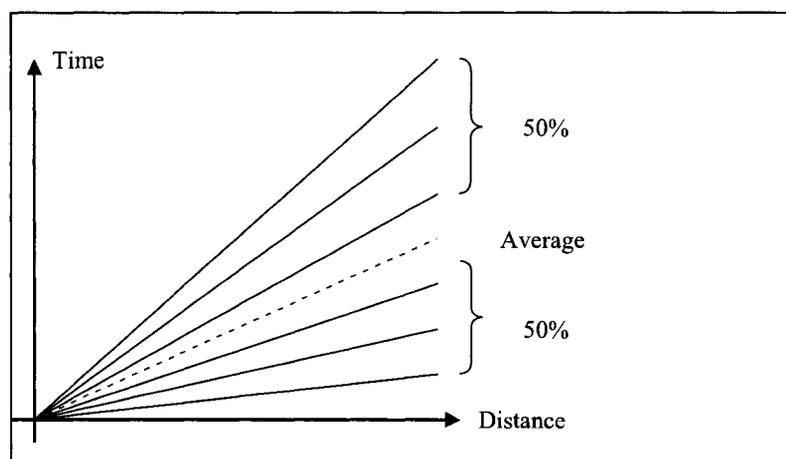


Figure 8.38: Late Car Travel Times are Compensated by Early Car Travel Times

In case of transit, the 85th percentile transit travel time (or a similar percentile) becomes the normal scheduled travel time (or the time allocated to the trip) and individual trips will not be shorter. Buses that travel ahead of schedule are normally delayed at ‘time points’ to prevent early running. Furthermore, buses arriving earlier than the allocated trip time (e.g. the 85th percentile time) will have extra recovery time but will not be able to start their next trip earlier (Figure 8.39).

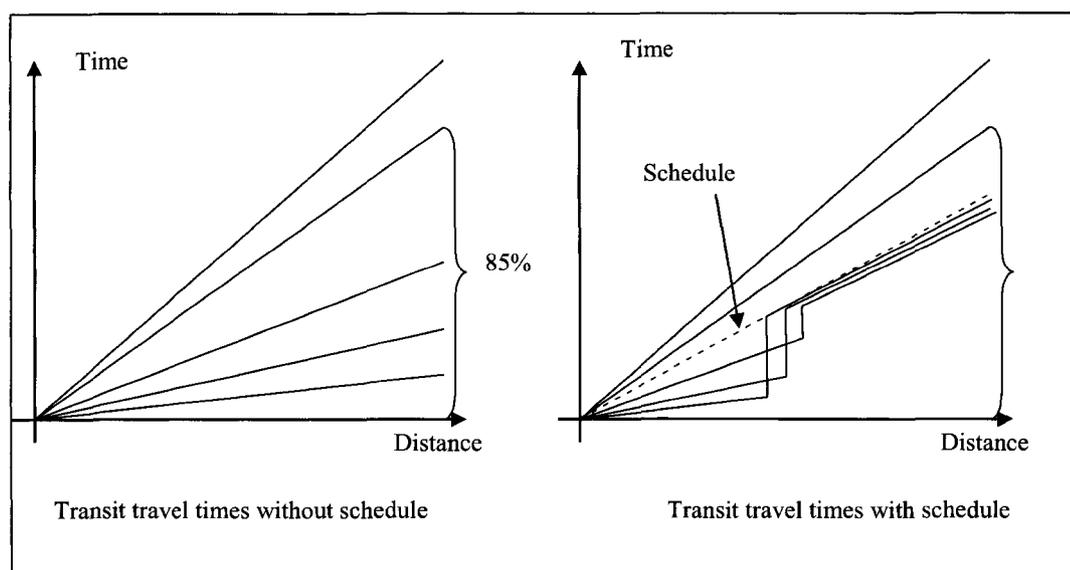


Figure 8.39: Early Buses Wait at Time Points or at the End of the Route

With the “Before” Scenario, bus travel time, for the observed section, is one and half times (1.5) longer than the car travel time. For scenarios A and B, transit travel time is much shorter than the car travel time (Table 8.22). The TCTTr indicates that the measure is likely to have some positive impact on TMS.

Table 8.22: Impact of Queue Management on the Relative Advantage of Transit Travel Times

	“Before”	Scenario A	Scenario B
Car travel time (mean)	420	443	432
Bus (85 th percentile travel time)	632	310	242
TCTTr	1.50	0.70	0.56

To estimate the impact of relative improvement of transit travel time on the TMS for the arterial, the definition of the corresponding origin-destination (OD) pair is required. As discussed in Section 5.4.4, a relatively simple procedure for determining the OD pair has to be researched and developed outside of the scope of this thesis. In the meantime, the TMS estimation will be carried out with hypothetical data only to demonstrate the estimation process.

Assume that the trip time between the origin and destination is 120 min and 60 min for the transit and car modes respectively for the “Before” scenario. Thus the TCTTr is 2. For Scenario B, the transit travel time is reduced by 242 sec which results in a TCTTr of 1.93 $((120-242/60)/60=1.93)$. The utility of transit for the relevant OD pair is calculated by Equation 5.44 and the corresponding TMS (P_{bus}) by Equation 5.23.

$$\Delta U_{bus} = 60 \times (1.93 - 2) \times \left(-0.026 + \frac{-0.05}{2 \times 25} \right) = 0.11$$

$$P_{bus}^{new} = \frac{0.15 \times e^{0.11}}{0.15 \times e^{0.11} + 0.85 \times e^0} = 0.16$$

Therefore, based on the discussed assumptions, the expected TMS for Scenario B is 16%, which represents a 1% increase.

8.10 Discussion and Conclusions

The partial queue jump is a very effective but rarely used bimodal queue management technique. The importance of the partial queue jump, with or without pre-signal, is that its implementation can be very wide spread. The design without a pre-signal is cheap to implement and operate.

The implementation of a pre-signal is expected to resolve the problem of lane violation and further improve the operating characteristics of the partial queue jump. Four pre-signal signal control algorithms were developed and tested in a simulation model environment. All four performed similarly, however there is difference between the four in terms of data requirements.

The partial queue jump technique was also demonstrated in a real-world example, which has been implemented in Ottawa. Further improvements to the implemented queue jump were also tested.

By utilizing the evaluation framework developed in Chapter 5 it is demonstrated that the partial queue jump with or without pre-signal improves the efficiency of the bimodal system in terms of the mean person travel time, it maintained the vehicular capacity, and it has a positive impact on transit travel times and TMS.

9. QUEUE RELOCATION TO THE UPSTREAM MAIN LINE APPROACH

9.1 Introduction

This Chapter presents the development and application of a queue management algorithm at an intersection that is upstream from the bottleneck intersection. The major challenges of developing queue management at upstream intersections is analysed and an algorithm is developed for a generic set-up and then applied to a simulated real world example. Performance measures are collected and discussed.

9.2 Description of Roadway Configuration

9.2.1 Physical Set-Up and Operational Characteristics before Queue Management

There are many situations where the implementation of a transit lane or queue jump at the bottleneck intersection approach or on the link leading to the bottleneck intersection is not feasible. In those situations it may be possible to implement a transit queue jump with queue management measures at the upstream signalized intersection. Using existing signals at upstream intersections for queue management is more complex than using pre-signals since they have already other function and there is much less flexibility in signal operation design.

A generic geometric set up is presented in Figure 9.1. The downstream intersection (B) represents the bottleneck on the presented roadway section. Assume that there is no opportunity to provide a transit queue jump at intersection B. However, there is opportunity to provide a transit queue jump in some form at the upstream intersection M.

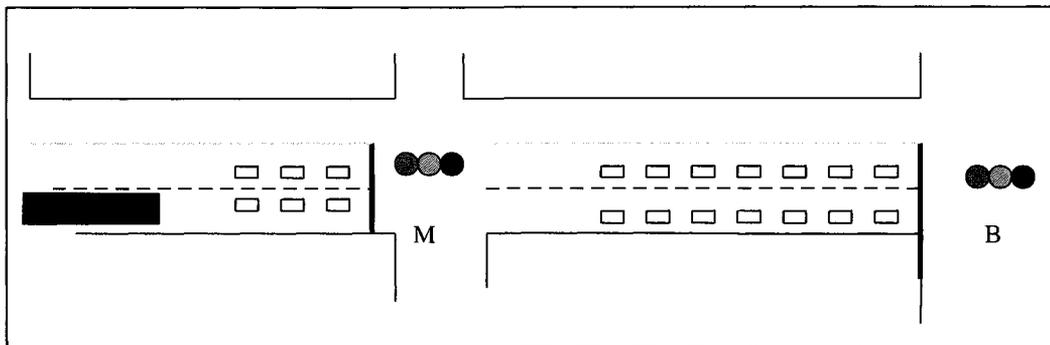


Figure 9.1: Elements of Upstream Metering

9.2.2 Physical Set-Up for Queue Management

The elements of the proposed queue jump set-up are presented in Figure 9.2. A transit queue jump lane is implemented at the approach to intersection M. Vehicles are counted by using check-in detectors for vehicles entering the restricted zone and check-out detectors for vehicles exiting the detection zone.

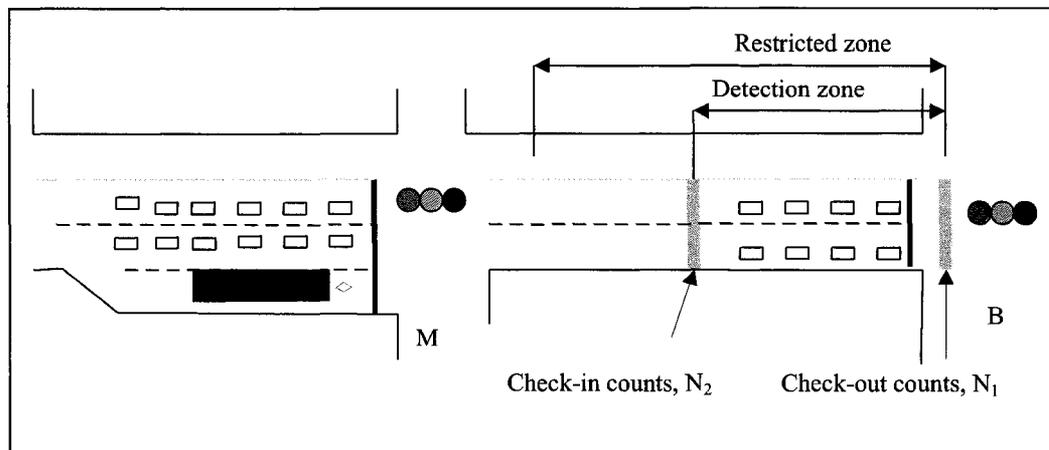


Figure 9.2: Physical Set-Up of the Proposed Queue Jump with Queue Management

9.3 Development of Queue Management Algorithm

The control objective is to manage the number of vehicles in the restricted zone which is defined somewhere between intersections B and M. By reducing the number of vehicles between intersection B and M the queue is relocated to the upstream intersection.

However, the queue jump at the upstream intersection allows the transit vehicle to pass the queue.

The control objective of the queue management algorithm is to adjust the green time at the upstream intersection, which is now called the metering intersection, in order to manage the number of vehicles in the restricted zone. The detection zone is part of the restricted zone and defines part of the road section where the number of vehicles are counted.

In the pre-signal control example (Chapter 8), the target number of vehicles in the detection zone was defined as the target queue length Q_T because at the time of taking the

sample measurement vehicles were stopped (except with “On-Off” Control). The target number of vehicles in the detection zone (NVDZ) is conceptually similar to target queue length, but the vehicles are not always stopped in the detection zone thus the term queue is confusing. Instead of using the terms target queue and measured queue, the terms NVDZ and target NVDZ are used.

The target NVDZ depends on the objectives of queue management. For example, the target NVDZ could be set to ensure that buses travel through intersection B during the cycle of arrival, or to prevent/minimize the blockage of auxiliary lanes (e.g. turning lanes), or to minimize local air and/or noise pollution at the restricted zone by limiting the number of vehicles waiting at intersection B.

Compared to the pre-signal, achieving a very accurate queue control is much more difficult. First, the distance between the metering and bottleneck intersections is longer thus the density of the arriving platoons within the detection zone changes very rapidly as the queue length at intersection B increases or decreases. Second, it is often difficult or undesirable to signalize every movement that leads into the restricted zone. For example right turns from the side street at the metering intersection are usually not controlled.

Another issue that needs to be resolved for queue management at upstream intersections is the distribution of upstream green times (splits). In traditional signal timing (developed for unsaturated conditions) splits are normally calculated to equalize the V/C ratios for all intersection approaches. This approach tends to equalize the relative queues at all

competing approaches. (The relative queue is defined as the measured queue divided by the maximum queue at capacity.) However, this equalization of relative queues may not provide a system optimum expressed by the minimum total person delay.

There are two alternative strategies for queue re-distribution. One alternative is to relocate all excessive vehicles to one approach only (Alternative B1 in Figure 9.3). The other alternative is to distribute the excessive vehicles to more than one approaches (Alternative B2 in Figure 9.3).

The selected queue management approach has to be based on the queue management strategy which is normally location and sometimes time specific. The general principle of bimodal traffic management is to re-locate vehicular queues to locations where they have the least impact on the movement of transit vehicles. For example, Alternative B1 would be selected if there is a transit queue jump lane on the eastbound approach. However, the re-distribution strategy would have to change to Alternative B at a time when the eastbound queue starts to block the queue jump lane access.

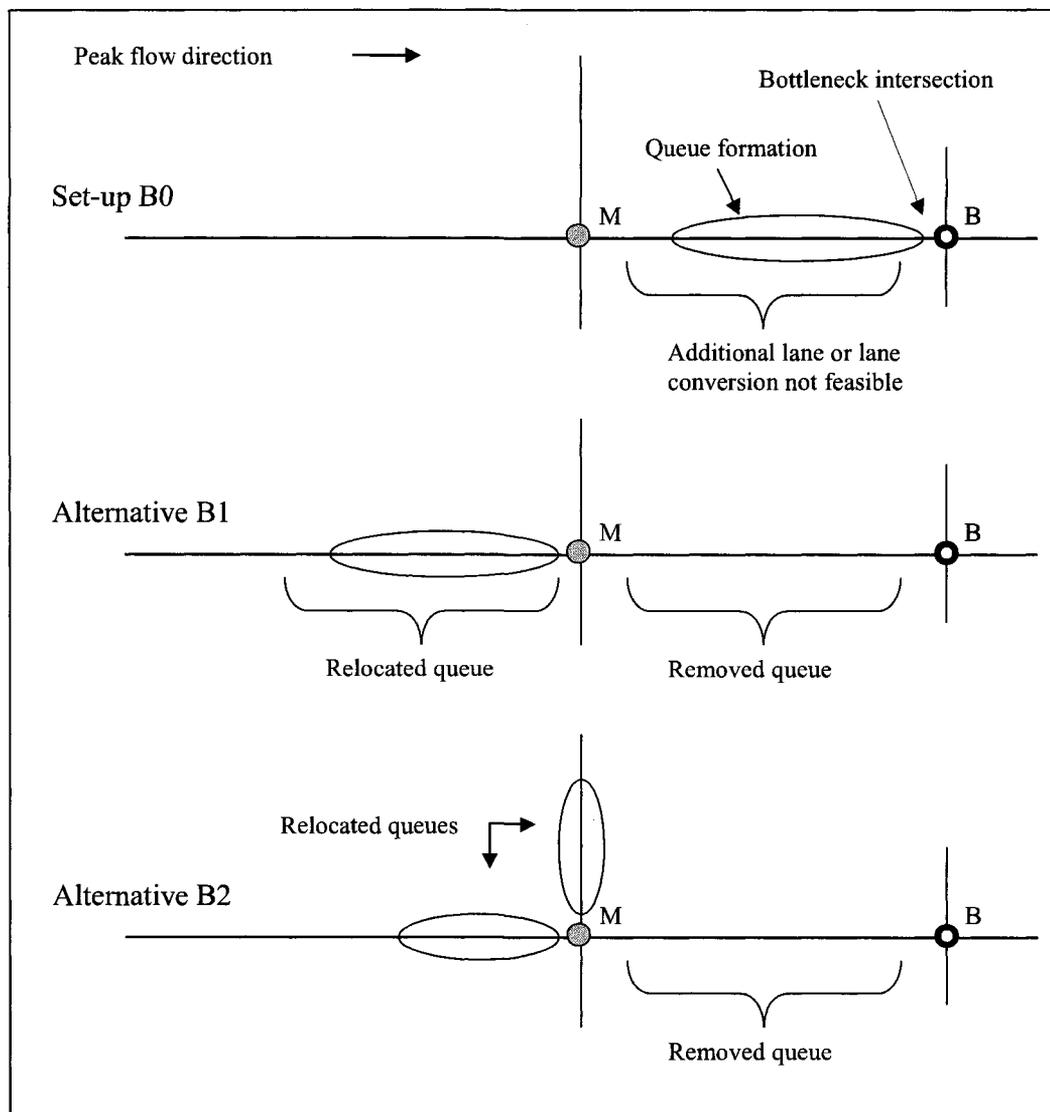


Figure 9.3: Queue Management with Queue Length Redistribution

The queue management control of upstream intersection M has two major parts:

Step 1: Calculation of the change of total discharge green time at intersection M

Step 2: Calculation of green splits at M.

Calculation of the change in the total discharge green time at intersection M

The initial signal timing at intersection M is calculated by using traditional methods. However, as the queue length at intersection B increases, the discharge green time at intersection M has to be reduced. The discharge green time at M consists of the green times of all signal groups that feed into the restricted zone. For example, for the signal timing plan described in Figure 9.4, the discharge green time is the sum of green times sg_2 and sg_7 .

Calculation of green splits at M

Before the calculation of green splits could be developed, the introduction of new definitions is required. Figure 9.4 represents the signal timing sequence for intersection M. The signal timing has five stages, and each stage has two signal groups. Signal groups sg_2 and sg_7 with the corresponding Stage 1 and Stage 2 feed into the restricted zone. Stage 1 and Stage 2 are called the *metering stages*, because the objective of the metering strategy is to increase or decrease the number of vehicles controlled by these two stages. The others, Stage 3, 4, and 5 are called *compensating stages* because they are either expanded or shortened to maintain a constant cycle length.

The calculation of green splits at intersection M has two components:

- distribute the total (allowable) discharge green time (from Step 1) between the metering stages (e.g. Stage 1 and Stage 2); and
- distribute the remaining green time between the compensating stages.

If the objective of the queue management approach is to relocate the queue to the upstream main line approach (Alternative B1 in Figure 9.3), the green time adjustment

generated in Step 1 is applied to Stage 1 (i.e. sg_2) and the duration of Stage 2 (i.e. sg_7) is not impacted by the green time adjustment.

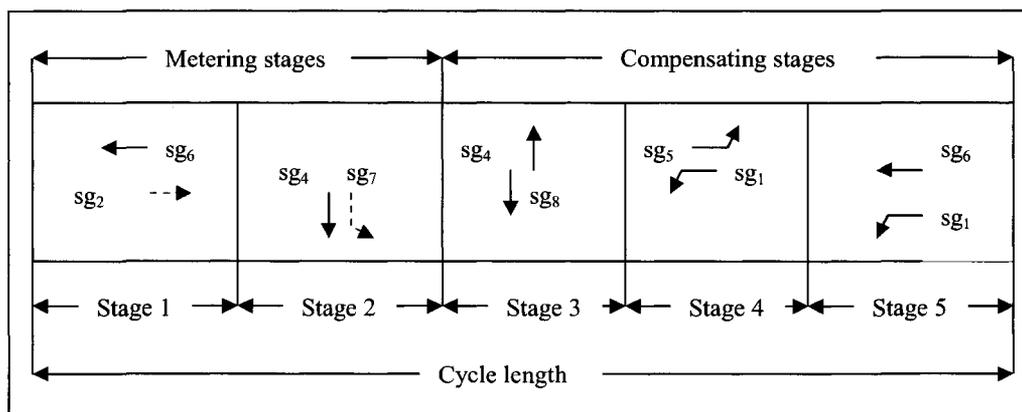


Figure 9.4: Signal Timing Sequence for Intersection M

In terms of the distribution of green time to compensating stages it is assumed that none of the compensating stages are congested, thus the allocation of the extra green time is not a major issue.

Therefore, if the outcome of Step 1 is a reduction of upstream green time, all the reduction will come from Stage 1 and the length of Stage 2 will not be affected. To compensate the signal cycle for the reduced length of Stage 1, a compensating stage, for example Stage 5 gets all the unneeded green time. At this point, Stage 5 is selected arbitrarily. However, the allocation of the extra green time can be based on specific criteria.

For the queue management example with the application of a pre-signal four different queue management algorithms were developed. The choice of algorithms in this case is however reduced.

The application of the “On-Off” and “Park View” control approaches is unfeasible. The length of the stages at the metering intersections must be known at the beginning of the cycle, and the “On-Off” and “Park View” control do not provide that information. There are at least two reasons why the length of the metering and compensating stages must be known at the beginning of the cycle:

- First, to display pedestrian phases properly, the control algorithm must know the duration of the entire stage at its start. Otherwise, the pedestrian walk display will have to be terminated immediately after timing down the minimum walk and the flashing don't walk intervals. Then the controller would rest in the don't walk display ready to terminate the stage. As a result pedestrian signal delay would increase.
- Second, to be able to allocate the extra time to the compensating stage, the amount of extra time has to be known at the beginning of that stage.

The Model based approach to this example has potentials but it is not pursued in this research for the following reasons:

- The primary objective of this research is to investigate the bimodal queue management approach, and the secondary objective is to investigate the potential of developing control algorithms based on human knowledge. Approaches based on human logic and heuristics may be more transparent and effective when applied to multiobjective problems such as the bimodal transportation management problem.

The Model based control was relatively simple for the pre-signal type queue

management approach, thus it was developed for comparison purposes. A Model based approach for this example would be significantly more complex.

- The Model based approach relies on precise measurements of all input values. As the area impacted by queue management grows, the number of locations where data has to be collected increases significantly. The availability of precise and reliable data from a large number of locations in a real-world environment cannot yet be assumed. Fuzzy Logic based control, in general, is better suited to problems where data is scarce in imprecise than model based control.

The controller algorithm is presented in Figure 9.5. This is a generic controller algorithm developed for this research. It is used in all other examples in this thesis with appropriate modifications. This algorithm contains only the basic elements of intersection control, therefore it is simpler than most controllers in the field. However, this simplicity does not represent an oversimplification. For example, the controller algorithm does not time down pedestrian intervals. However, the pedestrian intervals are built into the corresponding vehicle intervals and pre-set minimum and maximum green times. This is an important detail, because leaving pedestrian timing out of controller operation significantly (but unrealistically) increases the flexibility of signal timing. Keeping the controller algorithm as simple as possible in this stage of development also helped to minimize the potential for programming errors. Block 12 in the algorithm represents the subroutine used to calculate the metering and compensating stages. Further details are provided in Figure 9.6.

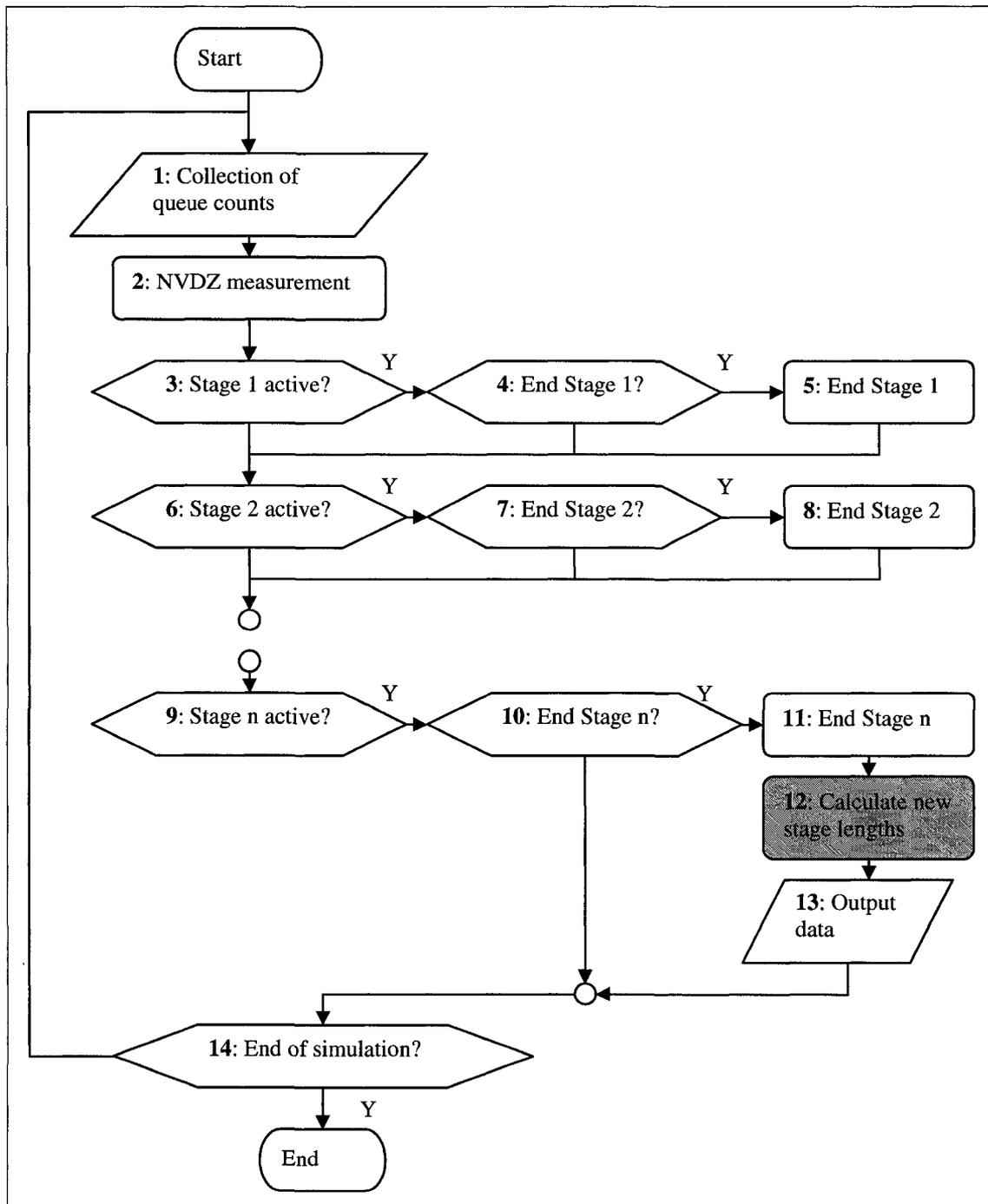


Figure 9.5: Controller Algorithm for the Metering Intersection

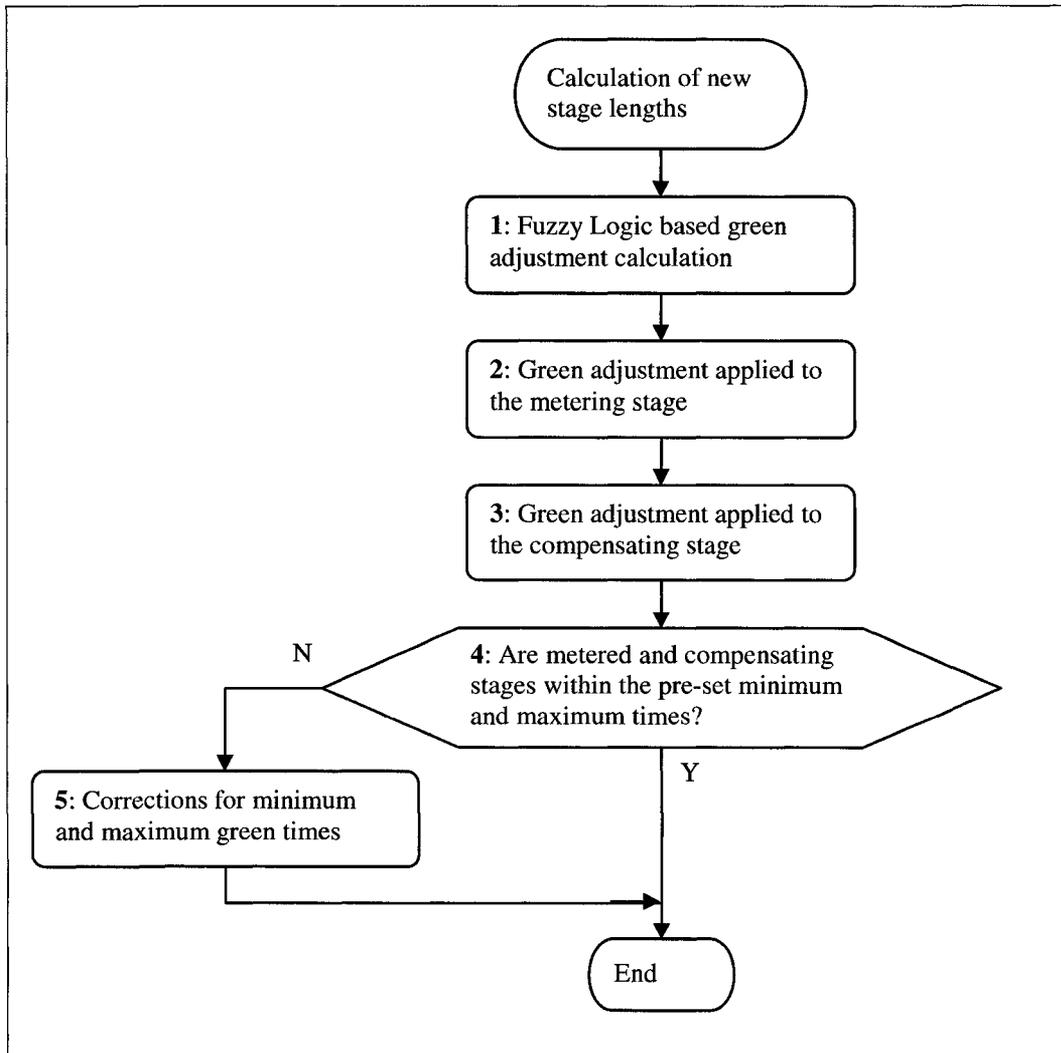


Figure 9.6: New Stage Length Calculations Subroutine

Block 1 in the algorithm includes the Fuzzy Logic based green adjustment calculation and it was been presented in Figure 8.15. In Blocks 2 and 3, the green times for the metering and compensating stages are adjusted, and in Block 4 and 5 the constraints to green times (minimum and maximum) are applied.

9.3.1 Tuning and Simulation

The Fuzzy Logic algorithm uses the normalized membership functions (Table 8.5, Table 8.6) and rule base (Table 8.8) developed for the pre-signal example. The Takagi-Sugeno type inference system is used with singleton output membership functions.

Although the FIS applied for this problem is identical to the FIS developed for the pre-signal, the system has to be re-tuned for the new geometry. The tuning process has to be repeated for every new geometric set-up.

The role of tuning is similar as for the pre-signal example (Figure 8.18). The only difference is that the objective of tuning is to maximize accuracy without special attention given to responsiveness because the target NVDZ is set to be a constant.

The tuning parameters include the scaling factor for the error sf_e , scaling factor for the change of error $sf_{\Delta e}$, and the output scaling factor sf_R (Section 6.5.2.1).

Tuning was carried out by trial and error. It was found that tuning was somewhat simpler than with pre-signals because the target NVDZ did not change during the simulation period.

The outputs were assessed based on the variability of the upstream discharging green time $\sigma(\text{green})$ and the difference between the target queue length and the measured queue lengths expressed as the root mean square error (RMSE).

The tuned parameters are presented in Table 9.1 along with the corresponding $\sigma(\text{green})$ and RMSE. The output NVDZ values, and the corresponding green times during a four hour simulation period are presented in Figure 9.7.

Table 9.1: Tuned Parameters and Control Performance with $Q_T=15$

Max green adjustment	Input scaling factor	Output scaling factor	$\sigma(\text{Green})$	RMSE
dn=7	$sf_e=0.5$ $sf_{\Delta e}=2$	$sf_R=0.2$ (-7,-6,0,6,7)	1.664	2.189

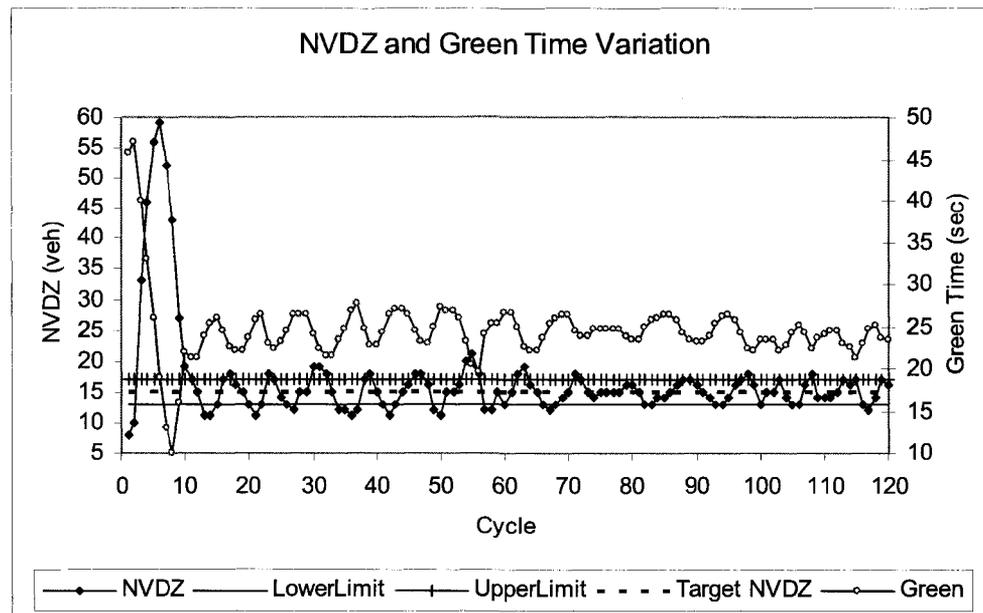


Figure 9.7: Control Output with $Q_T=15$

The settling period at the beginning of the simulation is rather long but that is not an issue. To eliminate the impact of initial errors, performance measures were collected after the 10th cycle. The reason for the relatively long settling period is that the design objective did not include high responsiveness thus no attention was given to it during the tuning process. The initial green time was set to 45 sec which is relatively high compared

to the required green time which is around 24 seconds, but the maximum green time change in the algorithm was limited to 7 seconds.

After the initial settling period, the measured NVDZ is mostly within +/- 5 vehicles from the target NVDZ which was initially set to 15 vehicles. This is somewhat higher than the variation of the measured queue length at the pre-signal example, however, this was expected as discussed in Section 9.3.

To assess the performance of the tuned FIS for different target NDVZ, the simulation was repeated for NVDZ values of 30 and 45 vehicles. The reason for assessing the sensitivity to different NVDZ target values is the following: The controller timing implementation process has two stages. In the first stage, the FIS is tuned based in an office environment by simulation. In the second stage during field implementation, the controller timing is fine tuned by adjusting the target NVDZ value. This fine tuning could be done in the field only if the FIS operation is not too sensitive to changes in the target NVDZ values.

The simulation was repeated two more times with a target NVDZ value of 30 and 45 vehicles. Simulation results show that the control algorithm is not too sensitive to different values of the NVDZ (Table 9.2, Figure 9.8, Table 9.3, Figure 9.9).

Table 9.2: Tuned Parameters and Control Performance with $Q_T=30$

Max green adjustment	Input scaling factor	Output scaling factor	$\sigma(\text{Green})$	RMSE
dn=7	sf _e =0.5 sf _{Δ_e} =2	sf _R =0.2 (-7,-6,0,6,7)	1.614	2.429

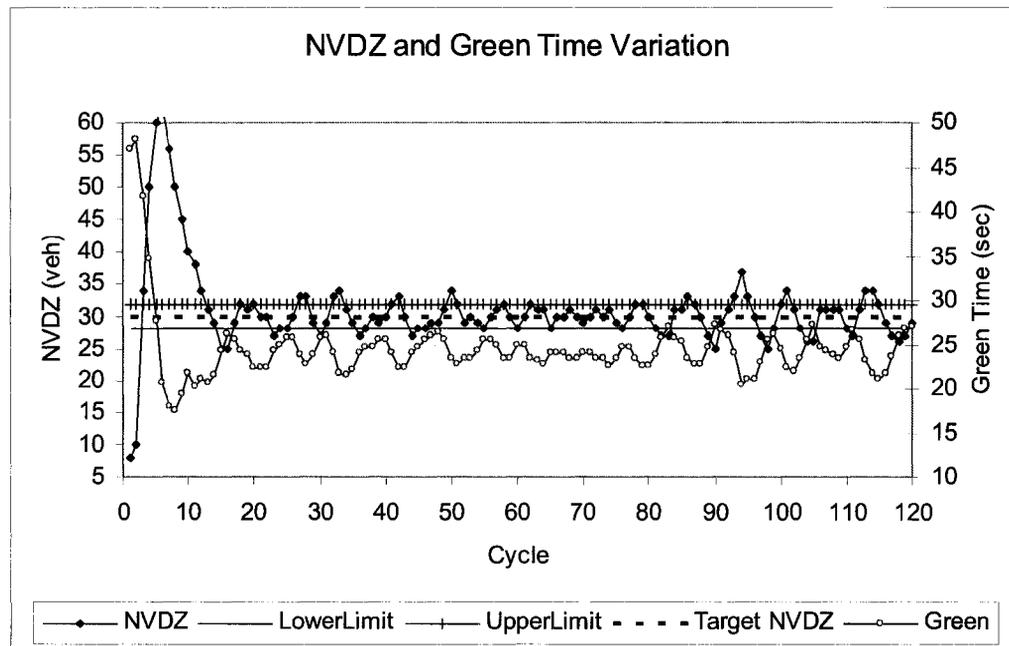


Figure 9.8: Control Output with $Q_T=30$

Table 9.3: Tuned Parameters and Control Performance with $Q_T=40$

Max green adjustment	Input scaling factor	Output scaling factor	$\sigma(\text{Green})$	RMSE
dn=7	$sf_c=0.5$ $sf_{\Delta c}=2$	$sf_R=0.2$ (-7,-6,0,6,7)	1.351	1.977

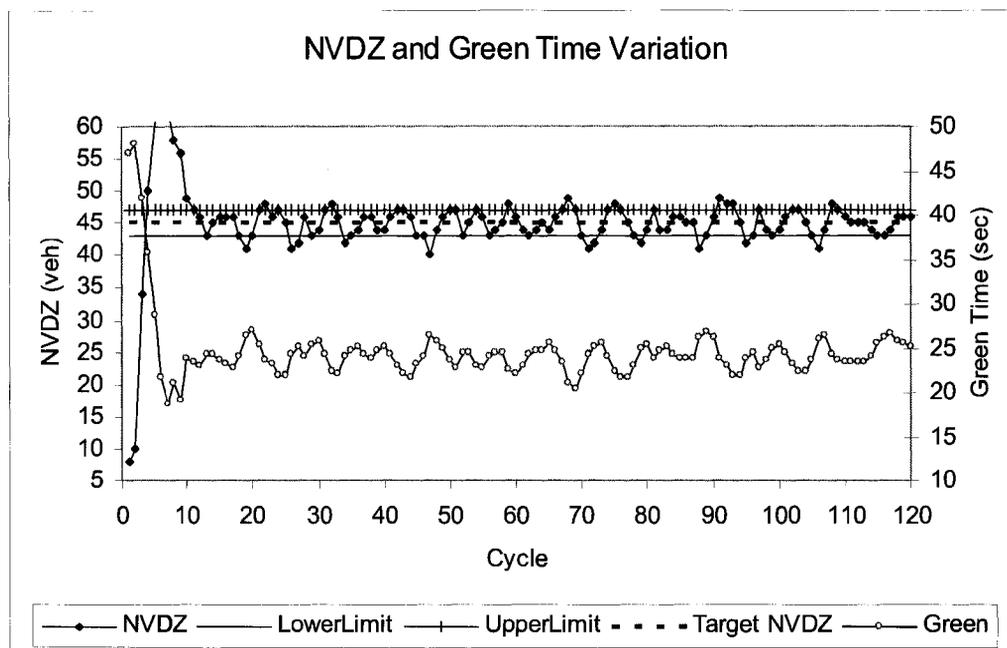


Figure 9.9: Control Output with $Q_T=45$

Thus, after the FIS has been tuned for a particular geometrical set-up in an office environment, adjustments to the target NVDZ can be carried out based on field observations.

9.4 Simulation of a Real-World Example: Woodroffe Avenue Corridor

9.4.1 Existing Corridor and Operations

Woodroffe Avenue is a major North-South arterial roadway connecting Barrhaven, a suburb with a population of approximately 70 000, with the two major East-West arterials (Hunt Club Road and Baseline Road) and a freeway, Hwy 417 (Figure 9.10). Woodroffe

Avenue, South of Hwy 417, has distinct northbound and southbound peak flows during the morning and afternoon peak hours respectively.

The general cross section has two mixed flow traffic lanes per direction plus the turning lanes at intersections. The South West Bus Rapid Transit (BRT) runs within the Woodroffe Avenue corridor. South from Hunt Club Road and North from Baseline Transitway Station, the BRT operates on a physically separated right-of-way, the Transitway. Between Hunt Club Road and Baseline Station, the BRT operates in curb side bus only lanes which are in addition to the two mixed flow lanes. In the southbound direction the curb side bus lane is continues between Baseline Transitway Station and Hunt Club Road, and vehicular congestion in the adjacent mixed flow lanes has minimal impact on transit travel times. In the northbound direction, however, the bus lane does not provide a continuous connection between the Transitway endings. There is an approximately 240 metre section between Meadowlands Drive and Algonquin Drive where buses travel in a mixed flow lane. Bus lanes were not provided along this 240 metre section because after leaving Meadowlands, buses leave the curb lane, and merge over to the median lane to make a left turn into Baseline Station at Algonquin (Figure 9.11). Transit travel times between Meadowlands Drive and Algonquin Drive have high variability because of the congestion in the mixed flow lanes.

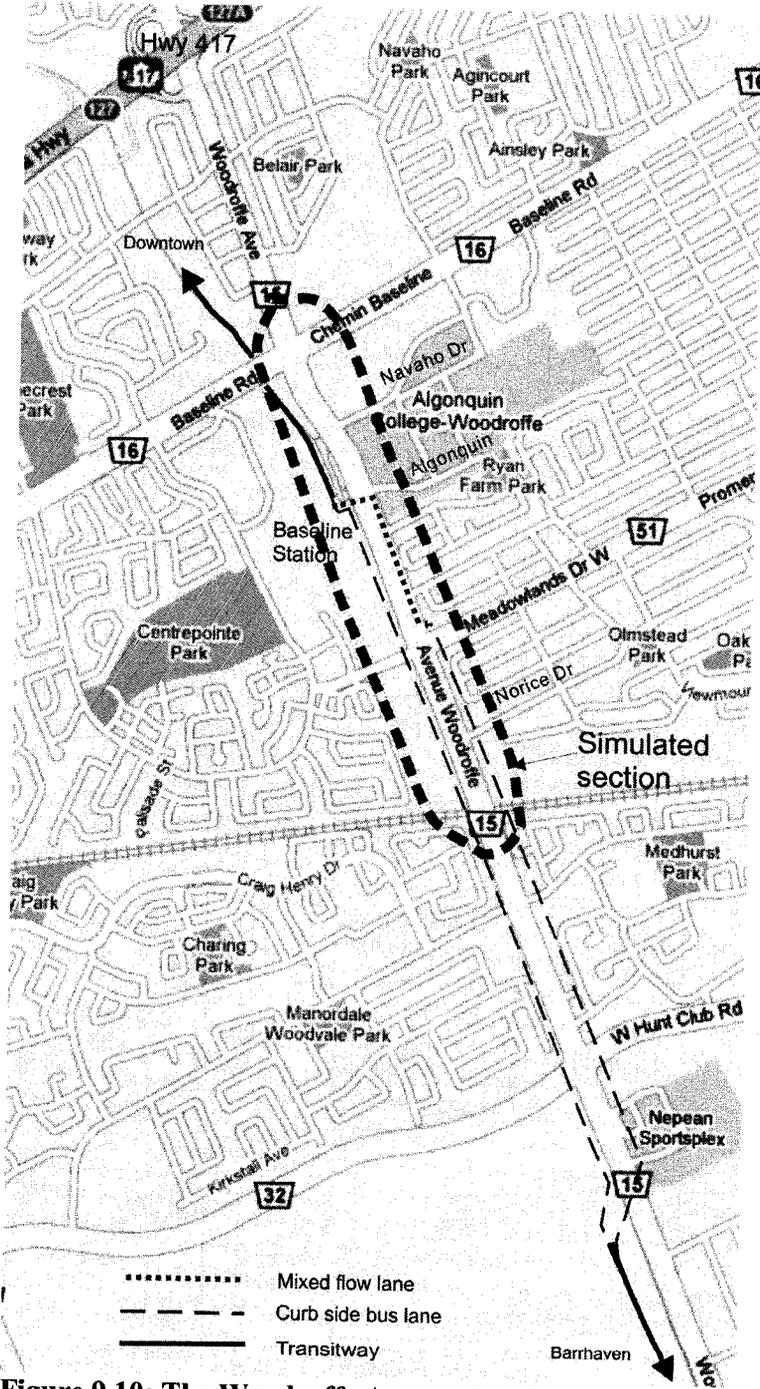


Figure 9.10: The Woodroffe Avenue Corridor

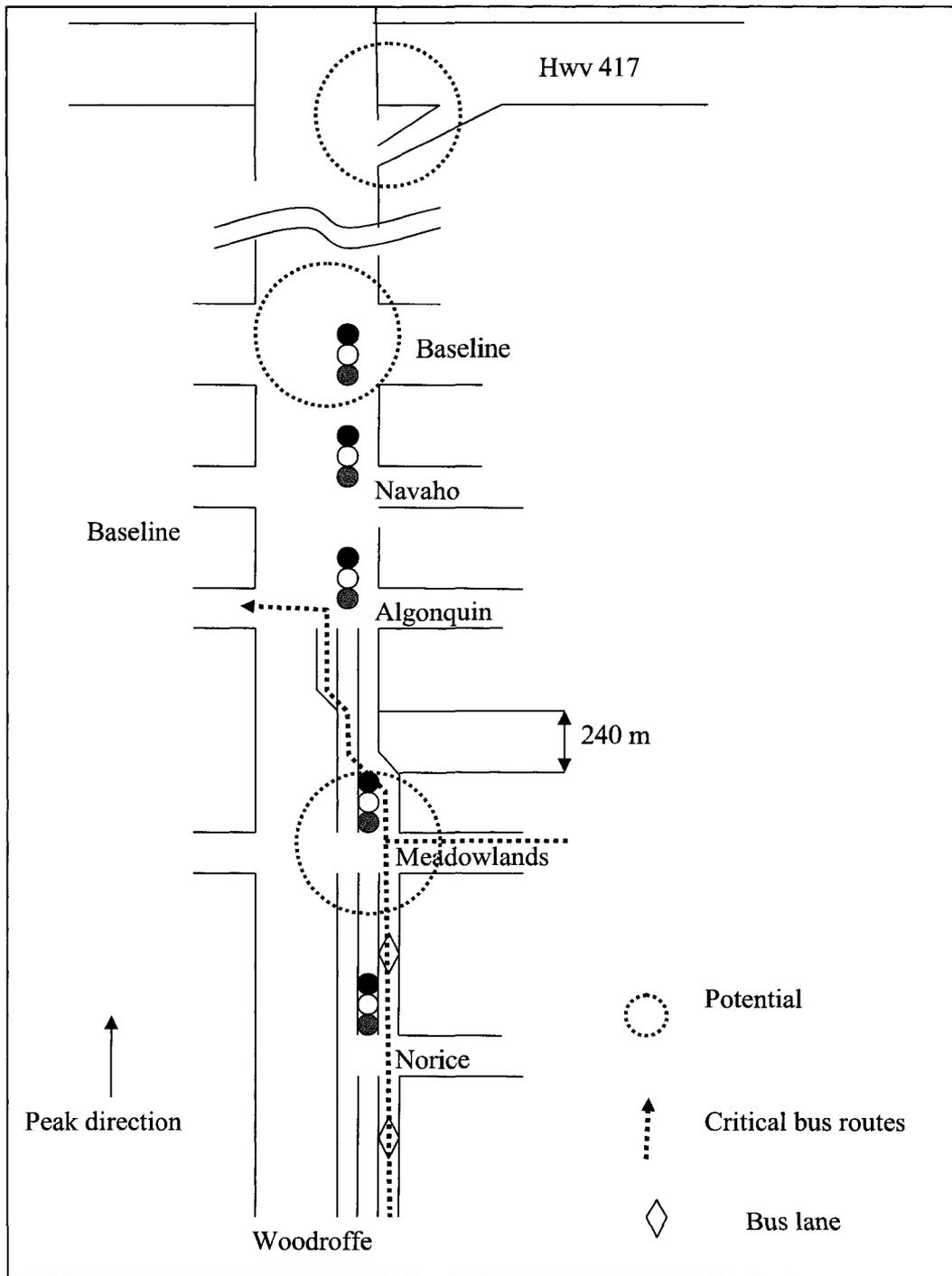


Figure 9.11: Schematic Lane Arrangement in the Critical Northbound Direction

The objective is to improve transit travel times by implementing queue management without significant negative impact on vehicular capacity or vehicular travel time.

However, before an effective queue management measure is defined, it is important to analyse and understand the locations of bottleneck(s) in the corridor. An analysis procedure for oversaturated conditions was developed in Section 7.3.2.2, and the procedure summary is repeated in Figure 9.12 for reference purposes.

The required background information about the corridor was obtained from different sources including:

- internet: roadway geometry was downloaded from aerial photos, transit volumes and routes were extracted from OC Transpo's web site;
- City of Ottawa database: traffic volume counts, signal operation parameters;
- site visits: general information about corridor operation; and
- City of Ottawa traffic operations staff: locations of recurring congestion.

Step 1: Volume Balancing

Volume balancing is presented in Appendix N.

Step 2 and Step 3: Determine signal operation and fine tune signal and operations parameters

Existing signal timing parameters obtained from the City of Ottawa were used as a starting point. The basic signal group sequencing was retained and stage splits were optimized manually for the balanced volumes. The manual optimization method, based

on equalizing phase splits was used because it is sufficiently accurate for the purpose of analysis and it is relatively simple and transparent.

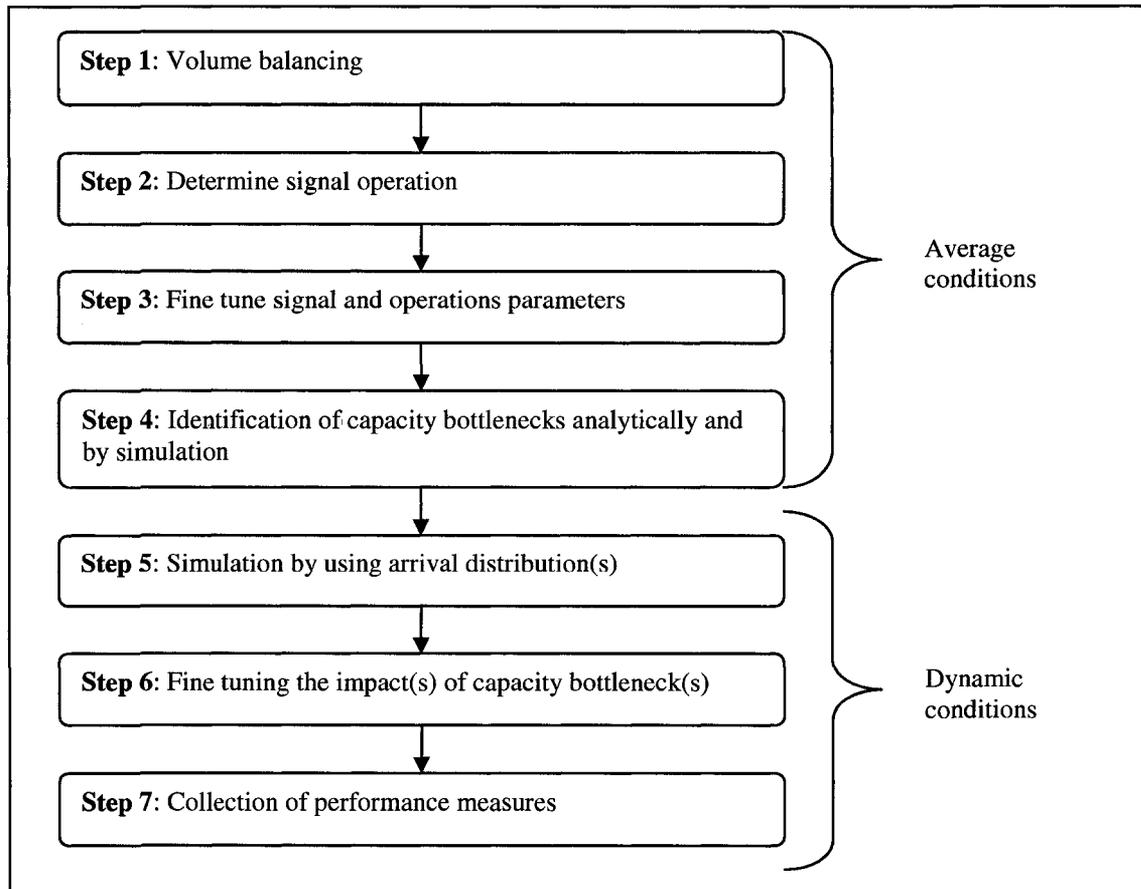


Figure 9.12: Analysis Process for Oversaturated Corridors (Copy of Figure 7.8)

To model real-world conditions as accurately as possible, special attention was given to the existing signal group sequencing. The Woodroffe Avenue Corridor is a designated Transit Priority Corridor (TMP, 2003) and a number of special transit signal priority measures have been implemented.

At the intersection of Woodroffe and Navaho, the green signal can be displayed twice during the signal cycle if triggered by eastbound buses (sg_4 and sg_T in Figure 9.13).

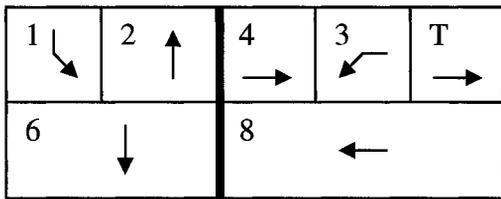


Figure 9.13: Phase Sequencing at Woodroffe Avenue and Navaho Drive

At Woodroffe and Algonquin the northbound left turn green can be also displayed twice in the cycle for buses entering into Baseline Station (sg_5 and sg_T in Figure 9.14).

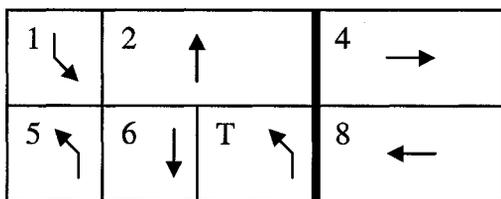


Figure 9.14: Phase Sequencing at Woodroffe Avenue and Algonquin

In the northbound direction at Woodroffe Avenue and Meadowlands Drive a short transit phase (sg_T in Figure 9.15) can be displayed before the opposing southbound left turning phase.

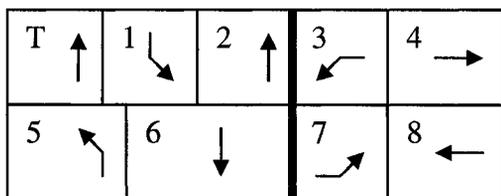


Figure 9.15: Phase Sequencing at Woodroffe Avenue and Meadowlands Drive

All transit signal groups are actuated, i.e. they are displayed only when a bus is detected on the intersection approach. The impacts of all these transit signal priority measures were included in the corridor operations simulation except the inserted transit signal group at the intersection of Woodroffe Avenue and Meadowlands Drive. The inserted transit signal group (sg_T on Figure 9.15) at Woodroffe Avenue and Meadowlands Drive

was not included in the analysis and simulation because the transit signal is displayed relatively infrequently. Bus travel between Norice Drive and Meadowlands Drive is not impeded by congestion and as a result of signal coordination most buses manage to enter the intersection during the regular northbound green phase.

More details on the signal timing process and signal timing parameters used in the simulation are provided in Appendix P.

Step 4: Identification of Capacity Bottlenecks by Simulation and Analytically

One of the critical steps in the analysis of oversaturated conditions is the identification of capacity bottlenecks. The standard approach to identify bottlenecks is through the traditional V/C calculation defined by Equation 9.1 and the results for the main intersections in the corridor are provided in Table 9.4.(V/C calculations including all input values are summarized in Appendix P.)

$$X = \frac{V}{n \times s \times \frac{g}{cyc}} \quad 9.1$$

X- degree of saturation (V/C ratio)

V – volume (veh/hr)

s – saturation flow rate (veh/hr/lane)

g- duration of effective green time (sec)

cyc – cycle length (sec)

n – number of lanes

Table 9.4: V/C Ratios at Major Intersections

	Meadowlands	Algonquin	Baseline
V/C ratio for NB ST	0.88	0.70	~0.88*

*10 % lane utilization of the curb lane is assumed based on field observations

Based on the traditional V/C calculation, the bottleneck intersection in the northbound direction is at Meadowlands Drive and/or Baseline Road because the V/C ratios at both intersections are the highest.

It must be noted that these numbers are only estimates since the inputs are highly variable. For example the saturation flow rate is impacted by weather conditions. A saturation flow rate of 1900 veh/hr was used in the calculations, because that rate has been measured in the field during the morning peak period on one sunny day. With a saturation flow rate of 1800 veh/hr, the V/C ratio at Meadowlands Drive and Baseline Road would increase to 0.93, very close to unity. Volume variability from one day to another is typically between 15% and 25 %. An increase of volume of 15 % at Meadowlands Drive increases the V/C ratio to unity.

Furthermore, the northbound approach to Woodroffe Avenue and Baseline Road intersection has three straight through lanes, but the curb lane ends about 90 metres after the intersection thus it is used less than the other two lanes. The utilisation of this lane is not constant, it increases with congestion. However, as the utilization of the curb lane increases, at one point the merging movement at the far-side of the intersection starts to negatively impact the saturation flow rate at the approach.

Nevertheless, the underutilized curb lane has some potential to provide some additional capacity at the approach as volumes increase. As a result, as volumes increase in the corridor (for example during the peak of the analysis period) the bottleneck is more likely to start to form at the intersection of Woodroffe Avenue and Meadowlands Drive.

This does not mean the intersection of Woodroffe Avenue and Baseline Road cannot become a bottleneck ($V/C=1$) in the corridor but it is unlikely that under normal conditions, based on the volumes used in the analysis so far, the northbound queues at Woodroffe Avenue and Baseline Road would reach the intersection of Woodroffe Avenue and Meadowlands Drive.

The above calculations and interpretations were confirmed by simulation. Based on the lane arrangement data, mean volumes, and signal timing parameters, traffic operation was simulated and visually checked to confirm operation in the field during “mean” conditions. The verification included the comparison of the number of vehicles entering the intersection at critical approaches in the field and by simulation.

Step 5: Simulation using arrival distributions

Traffic counts provided by the City of Ottawa included 15 minute volumes which served as the basis for developing arrival profiles (Appendix R). The final arrival profile used in the simulation (Table 9.5 and Figure 9.16) is called the Assumed Distribution and it was determined through an iterative process which is further explained in Step 6.

The first interval (Table 9.5) is a start-up interval during which no data is collected for evaluation purposes. The purpose of this interval is to populate the network with vehicles. The total arrival volume during intervals 2 to 5 (inclusive) correspond to the mean hourly volume. Arrival volumes gradually drop after the fifth period to zero to eliminate the impact of any vehicles that remained in the system at the end of the evaluation period.

Table 9.5: Arrival Volume Distribution at the Most Upstream External Link

Time period (sec)	Volume Share (%)	Volume (veh/period)	Volume (veh/hr)
0-720			1800
720-1440	19	432	2160
1440-2160	20.5	466	2331
2160-2880	21	478	2388
2880-3600	20.5	466	2331
3600-4320	19	432	2160
4320-5040	n/a	n/a	1800
5040-5760	n/a	n/a	1200
5760-6480	n/a	n/a	600
6480-7200	n/a	n/a	0
Total:	100	2274	

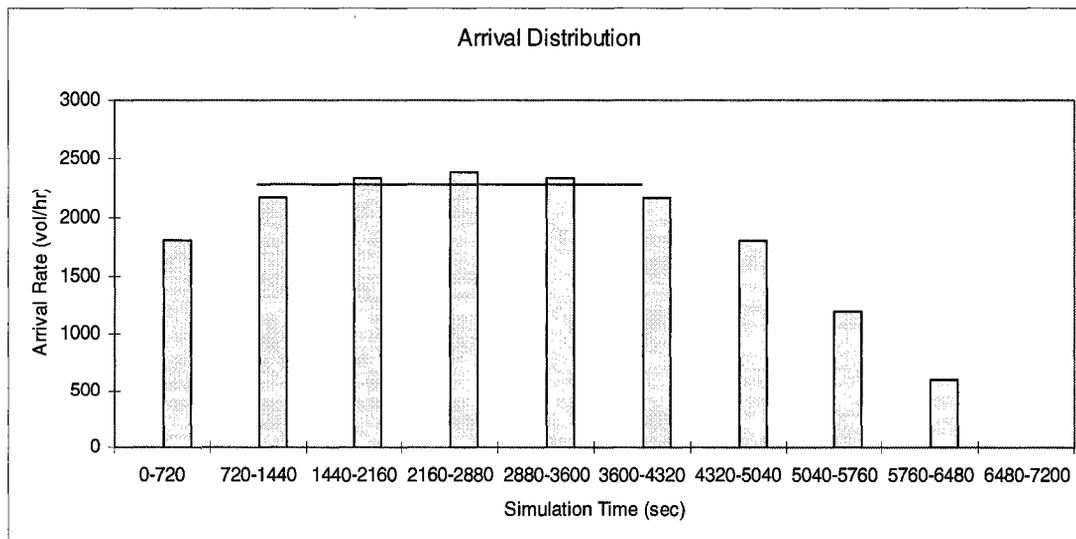


Figure 9.16: Assumed Distribution of Arrivals

Step 6: Fine tuning the impact(s) of capacity bottlenecks

If the results of the V/C calculations provided in Step 4 and Table 9.4 were a true and consistent indication of traffic conditions in the field, the downstream intersection from Meadowlands would rarely experience congestion. The metering effect of the intersection of Woodroffe Avenue and Meadowlands Drive would prevent the development of oversaturated conditions at the adjacent northbound intersections. This is, however, not always the case.

Based on empirical evidence, it is known that when Hwy 417 becomes congested the northbound to eastbound ramp restricts traffic flow in the Woodroffe Avenue corridor more than the upstream intersection of Woodroffe/Baseline or Woodroffe/Meadowlands. Queues often start to develop at the on-ramp, and the end of the queue can reach Baseline Road, Algonquin, Meadowlands Drive and even further upstream.

Based on the above analysis, it can be concluded that there are at least two potential bottlenecks in the corridor, one at the intersection of Woodroffe Avenue and Meadowlands Drive and the other at the on-ramp to Hwy 417 – which is actually outside of the original scope of analysis. Field observations over a number of years confirm the existence of multiple bottlenecks. Sometimes queues form at Woodroffe Avenue and Meadowlands Drive while downstream from Meadowlands operating conditions are much better, and there are times when queues are present in the entire corridor all the way up to the Hwy 417 on-ramp.

The repercussions of the presence of multiple bottlenecks are two-fold:

- First, when simulating corridor operation, one has to decide which bottleneck's impact will be analysed.
- Second, when designing queue management measures, one has to know which bottleneck's impact will be controlled.

The selection of bottleneck for the purpose of simulation and queue management is trivial in this case. The problem of the bottleneck at Woodroffe Avenue and Meadowlands Drive from the bimodal transportation perspective has been resolved a number of years ago by creating a queue jump lane from Norice Drive to Meadowlands Drive first, and then by extending the queue jump lane into a continuous bus-lane. This leaves, the impacts of the other bottleneck to be analysed and resolved from the bimodal transportation perspective.

The direct simulation of bottlenecks in oversaturated networks is not always practical. For example, to simulate the operation of the Woodroffe/Hwy 417 on-ramp bottleneck, one would have to simulate the operation on Hwy 417. However, to show the congested conditions on Hwy 417, one would have to model downstream operation on Hwy 417 including all the off-ramps and all the signalized intersections downstream of the off ramps that impact operation. This is obviously not practical, and the scope of simulation must be reduced to a reasonable level.

To resolve the problem of scope-creeping, the downstream bottleneck at the Hwy 417 on ramp was simulated indirectly by reducing the speed of vehicles in the northbound

direction at the far-side of the intersection of Woodroffe Avenue and Baseline Road. The slow moving traffic reduces the discharge flow for northbound traffic at the intersection of Woodroffe Avenue and Baseline Road.

The operation of an intersection or corridor in oversaturated conditions is very sensitive to input parameters such as volume (total and distribution), signal timing, and the saturation flow rate. The sensitivity is reflected in large variations of queues and delays.

To ensure that the queues always remain within the simulated network, fine tuning some of the critical parameters is necessary. Fine tuning was achieved by slightly modifying the arrival profiles to ensure that duration of oversaturated conditions correspond to the empirical observations. Furthermore, the saturation flow rate at Woodroffe Avenue and Baseline Road for northbound traffic was tuned indirectly by changing the departure speeds from the intersection. The main tuning objective was to keep queue formations for the evaluated movements within the simulated network.

The above tuning process is different from “traditional” tuning. The objective of traditional tuning is to adjust simulation parameters so that real-world conditions are modelled as realistically as possible. For simulating the impacts of control in oversaturated conditions, the above objective is still valid in general terms but the objective of keeping all impacts within the corridor is dominating.

Step 7: Data collection for performance measures

When selecting the data collection points (for defining travel time sections) for evaluating queue management measures it is important to know the exact impacts of the queue management algorithm. The queue management algorithm in the Woodroffe corridor changes the duration of the green signals for the northbound and southbound straight through, and northbound and southbound left turn movements (sg₂, sg₆, sg₅, and sg₁ respectively) (Section 9.3). Therefore, travel times impacted by all these signal timing changes have to be collected to be able to comprehensively evaluate the impacts of the implemented queue management measure. The data collection points for car and bus travel times are defined in Figure 9.18 and the descriptions of travel sections are defined in (Table 9.8).

9.4.2 Proposed Geometric Set-Up and Operation

The queue between Meadowlands and Algonquin has a significant negative impact on transit travel time and variability. The objective of queue management is to minimize the impact of congestion on transit travel times by reducing the length of queues between the intersections with Meadowlands and Algonquin.

The queue management algorithm was developed for this location to move part of the queue upstream from Meadowlands Drive where it has a smaller or no impact on transit vehicles. Normally, the second objective of the queue management algorithm is to maintain maximum flow (described as existing capacity) through the bottleneck

intersection. However, the bottleneck intersection at the Hwy 417 off-ramp, for which the queue management control is designed, is significantly downstream from the roadway section where the queues will be managed. As a result there is no need to focus on maintaining the capacity of the bottleneck intersections because the need for queue management between Meadowlands and Algonquin will cease before the queue at the downstream bottleneck intersection will discharge.

The geometric set-up for queue management control is presented in Figure 9.17. The biggest difference between this and previous set-ups (Figure 8.3, Figure 9.2) is in the placement of the detection zone which was moved downstream from the intersection of Woodroffe and Algonquin. The detection zone is actually outside of the restricted zone. The detection zone location selection was based on numerous trials. By placing it far-side of the intersection the responsiveness of the control was improved. The metering intersection (Woodroffe Avenue and Meadowlands Drive) is relatively far from the intersection where the queue is controlled. Thus, there is significant delay between the queue formation at the intersection of Woodroffe/Algonquin and the control/metering action at the upstream intersection. By moving the detection zone downstream of the intersection of Woodroffe/Algonquin, this reaction delay is reduced.

The generic queue management algorithm (Figure 9.5) and the stage length calculation subroutine (Figure 9.6), developed and presented earlier in this chapter, were used. However, since the geometric set-up differs from the set-up used for tuning, the FIS had to be re-tuned.

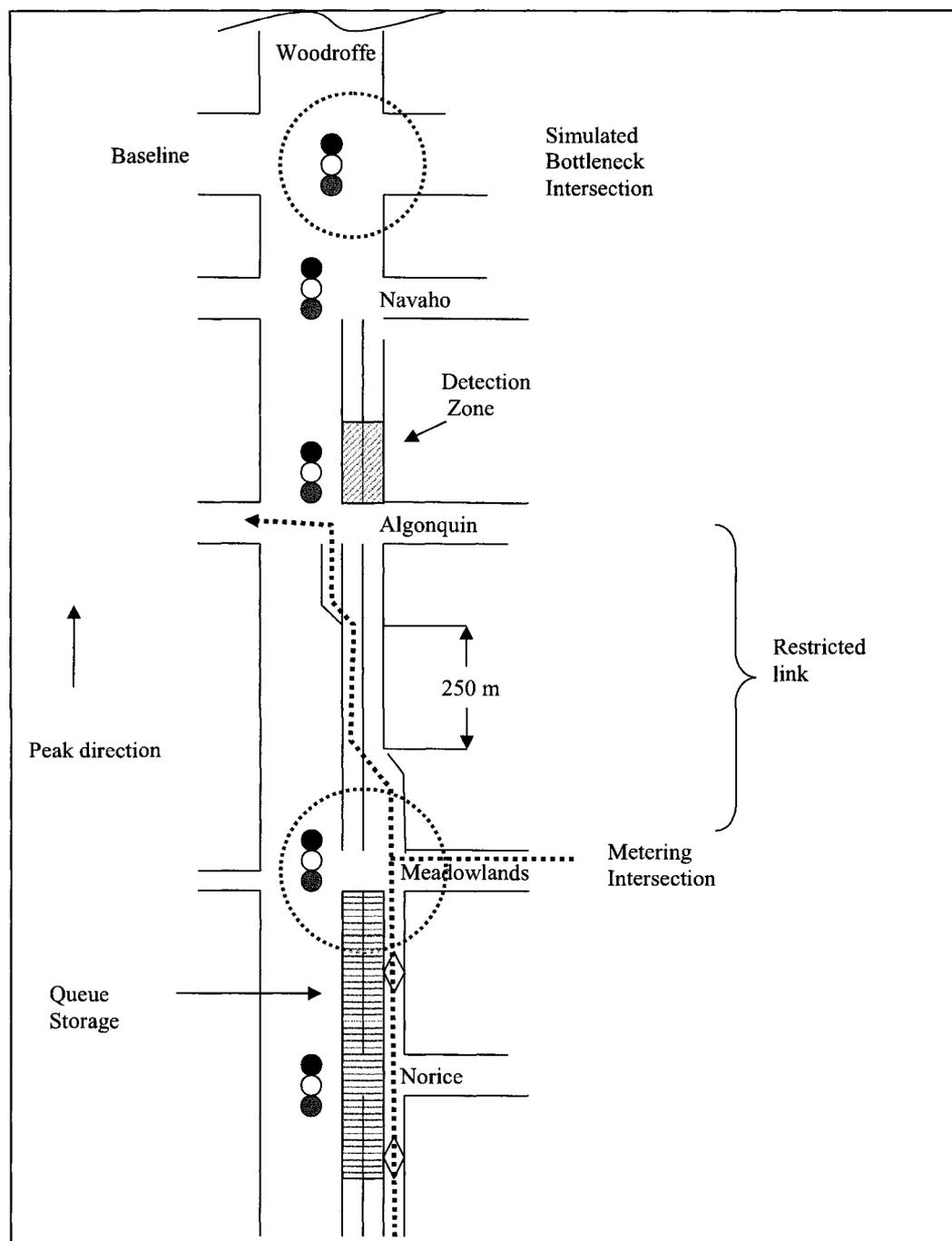


Figure 9.17: Geometric Set-up for Queue Management

9.4.3 Simulation Results

The evaluation framework proposed in Chapter 5 is followed with some changes. Since the queue management algorithm impacts many travel patterns within the simulated area, there is a special focus on each travel pattern. In addition to the vehicle travel time, person travel time, transit system operations perspectives, two more aspects of the queue management operation are examined. The first is sensitivity of the control algorithm to different levels of congestion, and the second is the impact on signal controller operation. Simulation results impacting transportation policy/planning objectives are not discussed here.

9.4.3.1 Vehicle Travel Time and Person Travel Time Perspectives

Overall impact on travel times

Vehicle travel times were collected for the entire simulation period (7200 sec) and the calculated means are presented in Table 9.6. If all vehicles are considered (cars and buses) the mean travel time of 269.0 seconds increased to 277.6 sec with the queue management control. However, the increase is only 8.6 sec (3.2%) and it is statistically not significant at the 95% confidence level.

Table 9.6: The Impact of Queue Management on Mean Vehicle Delay

		Mean Vehicle Travel Times (sec)		
		No QM (sec)	With QM (sec)	Difference (sec)
Car	Mean	270	280	10 (3.5%)
	σ	33	36	
Bus	Mean	228	213	-15 (-6.4%)**
	σ	8	2	
Car+Bus (Total)	Mean	269	278	9 (3.2%)
	σ	32	34	

** Significant at the 99% confidence level

Total person delay is considered to be a more appropriate measure of the effectiveness of multi-modal transportation; therefore this MoP was calculated and is presented in Table 9.7. The calculation of person delay was based on vehicle delays for the two modes and assumed vehicle occupancy values of 1.2 per/veh and 40 per/veh for the car and bus modes respectively.

With the queue management control, total person delay decreases from 841.7 veh-hours to 828.3 person-hours which represent a statistically insignificant reduction of 1.6 %.

Table 9.7: The Impact of Queue Management on Total Person Travel Time

		Total Person Travel Times (pers-hr)		
		No QM (sec)	With QM (sec)	Difference (sec)
Car+Bus (Total)	Mean	841.7	828.3	-13.4 (-1.6%)

** Significant at the 99% confidence level

The general expectation was that the mean vehicle travel time will stay the same or slightly increases but the total person travel time will decrease. This expectation has been

only partly fulfilled. One can say that there is evidence that the mean vehicle travel time has not changed, thus queue management did not make things worse. However, since the total person travel time change is statistically insignificant as well, one could also argue that the measure is ineffective, thus there is no reason to implement it.

To demonstrate the effectiveness of the simulated queue management measure, the analysis of aggregate measures such as the mean vehicle travel time and the total person delay collected for all users in the study area during the entire simulation period is not sufficient. The aggregate measure masks the important changes.

By gradually disaggregating the collected simulation data, it is possible to reveal the significant changes resulting from queue management. The first level of disaggregation is carried out by differentiating between car and transit travel times.

The objective of the queue management design was to improve transit travel times compared to car travel times without making car travel times significantly worse. The results summarized in Table 9.6 show that this objective has been achieved. Mean bus travel times for all routes decreased by approximately 15 seconds (from 227.8 seconds to 213.2 seconds) which is statistically significant at the 99% confidence level. At the same time, the mean car travel times for all vehicles in the simulated network increased by approximately 10 seconds (from 270.5sec to 280 sec) which is statistically insignificant at the 95% confidence level (Appendix S). Similarly, Table 9.7 shows that there is a

statistically insignificant increase in the person travel time for the car mode but there is a statistically significant decrease in the total person travel time for the transit mode.

Impact on different travel patterns (different travel sections)

Bimodal queue management relocates vehicular queues from locations where queues have negative impact on transit vehicle operation to locations where they have less impact. Therefore, it is expected that the impact of queue management is different for various travel segments within the simulated network. Travel segments are listed in the first column of Table 9.8 and the origin and destination points are depicted in Figure 9.18.

The section based travel time analysis shows that there has been a redistribution of travel times between different travel sections. For example, northbound travel times A to C, A to D increased significantly. Cars travelling northbound in this section are negatively impacted by the relocated queue from the section between Meadowlands and Algonquin to Norice Drive and Meadowlands Drive but do not benefit from the absence of queue between Meadowlands Drive and Algonquin. At the same time, vehicles entering the simulated network from Meadowlands and travelling north benefit from the absence of queue between Meadowlands and Algonquin (C to E, C to G, C to H, D to E, D to G, D to H). Vehicles travelling north in the Woodroffe corridor up to Algonquin have very similar travel times with and without queue management. The increased travel time in the relocated queue south of Meadowlands is compensated by the absence of queue delay between Meadowlands and Algonquin.

Table 9.8: Comparison of Travel Times at Specific Sections with no Queue Management and with Queue Management

Section	Mean Travel Time		Difference	
	Without QM	With QM	Absolute	Relative
A to C	119	167	48 **	40%
A to D	181	227	46 **	25%
A to E	239	262	23 *	10%
A to G	293	298	5	2%
A to H	392	395	2	1%
C to E	64	54	-11 **	-16%
C to G	109	84	-25 **	-23%
C to H	240	201	-39 **	-16%
D to E	202	158	-44 **	-22%
D to G	263	223	-40 **	-15%
D to H	417	354	-63 **	-15%
H to C	147	136	-11 **	-7%
H to B	130	133	3	2%
H to A	196	196	0	0%
A to F (Bus)	270	246	-24 **	-9%
C to F (Bus)	155	135	-20 **	-13%
F to A (Bus)	216	219	3 **	1%
F to C (Bus)	184	174	-10 **	-6%

* Significant at the 95% confidence level ** Significant at the 99 % confidence level

In the southbound direction car travel times did not change significantly except for the travel section from H to C. The reason for this reduction is that the southbound left turn stage is the compensating stage in the queue management algorithm thus this stage gets all the unused green time.

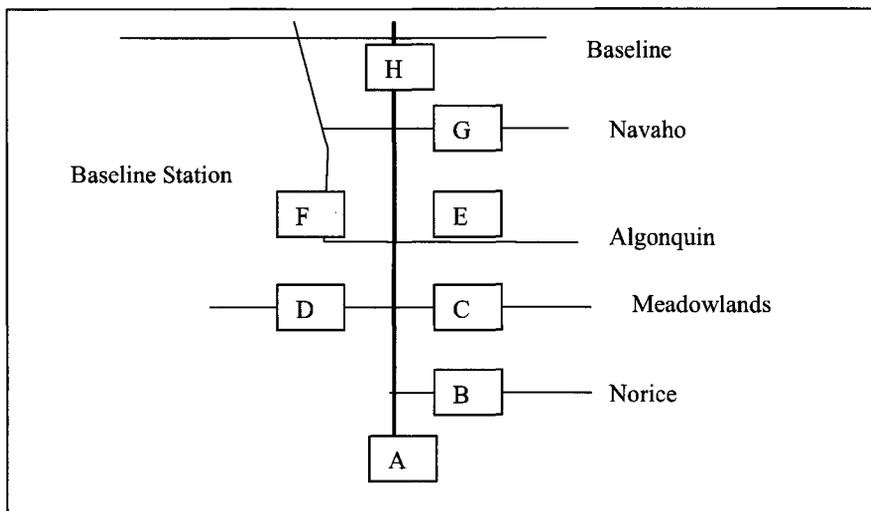


Figure 9.18: Definition of Travel Sections

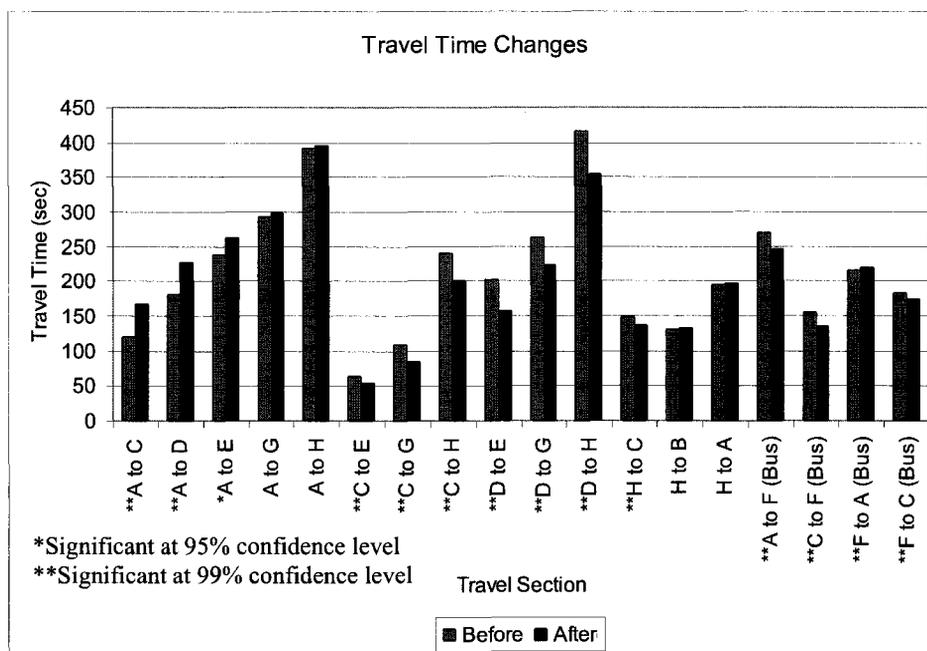


Figure 9.19: Comparison of Travel Times at Specific Sections without (“Before”) and with (“After) Queue Management

Northbound transit travel times decreased for sections A to F and C to F because buses are not delayed by the queue between Meadowlands and Algonquin. South bound transit travel times also decreased for section F to C for the same reasons as car travel times on this section, the southbound left turn green time at Woodroffe and Meadowlands is longer on mean because this stage is the compensating stage. However the southbound transit travel time from F to A increased. The increase is small, approximately 1 %, but statistically significant because transit travel time variability in the southbound direction is very small. The increase is due to the shorter mean southbound green time with queue management.

Impact on Car Travel Time Variability

Further desegregation of the mean travel time based on the 2 hr simulation period to 6 min periods reveals the variability of travel times within the observation period. Car travel time variability is depicted on Figure 9.20. From A to H, travel time is essentially the same during the entire observation period with and without queue management. As discussed earlier, the additional travel time between Norice Drive and Meadowlands Drive has been compensated by the lower travel time between Meadowlands and Algonquin.

However, the increase of car travel times is evident for sections A to C and A to D (Figure 9.20). For example, from A to C car travel times are higher by up to approximately 170 seconds with queue management. As discussed earlier, cars travelling from A to C and A to D are delayed by the relocated queue but do not benefit from the

uncongested section downstream from Meadowlands because they turn off Woodroffe at Meadowlands.

In contrast to cars that leave to corridors at Meadowlands Drive, vehicles that enter the corridor at Meadowlands Drive and travel northbound only benefit from the queue management. Travel times decrease for sections C to H and D to H. For example, from D to H, travel time decrease is up to approximately 180 seconds.

This is intentionally empty space

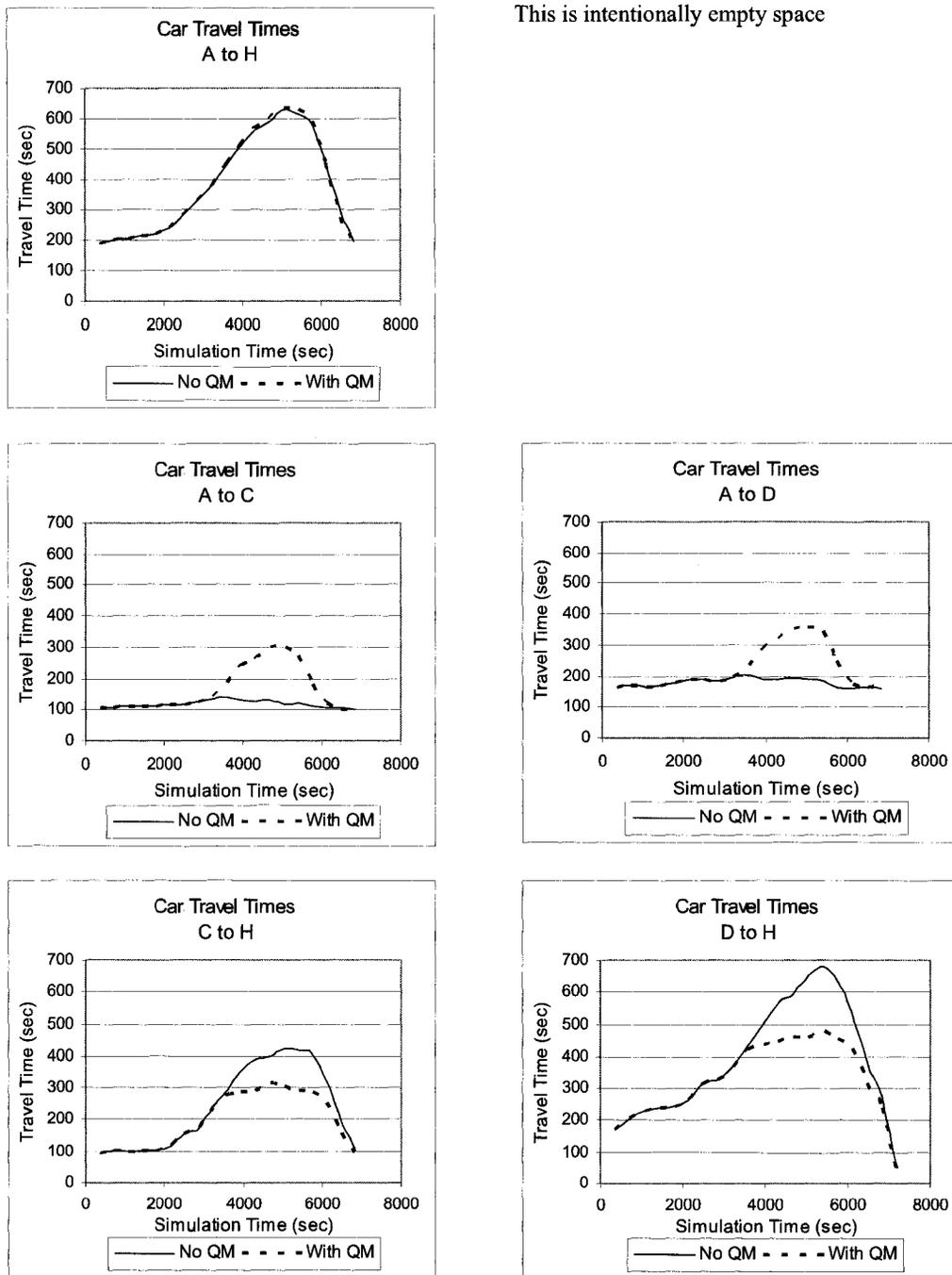


Figure 9.20: Comparison of Car Travel Times During the Simulation Period

9.4.3.2 Transit System Perspective

Impacts of Queue Management on Transit Travel Time Variability

The impact of travel time variability on the transit mode is much more important than on the car mode. From the perspective of individuals, unpredictable travel times are equally disruptive to both car and transit passengers. However, from the transit system operations perspective unpredictable travel times (e.g. travel times with high variability) have profound impact on the operation of a schedule based transit system. First the impacts of queue management on transit travel time variability is analysed and then the implications of transit travel time variability on transit system operations are explored.

Without queue management, the mean transit travel time from A to F calculated for 6 min periods varies from approximately 250 seconds at the beginning and at the end of the observation period to 350 seconds during the peak part of the observation period. With queue management, this variability has been mostly removed. A similar observation can be made for transit travel times from C to F.

For southbound transit travel times, the variability of transit travel times is much less because they are not impacted by vehicular queues. The differences between travel times with and without queue management are due to the changes to the duration of southbound green times at the intersection of Woodroffe Avenue and Meadowlands Drive. The southbound green time becomes shorter (metering stage) thus travel times from F to A increase slightly, and the green time for southbound left turns becomes longer (compensating stage) thus travel times from F to C decrease slightly.

The presentation in Figure 8.21 also demonstrates why the analysis of mean travel times during the entire observation period is not sufficient for oversaturated conditions. The mean transit travel time reduction for A to F during the 2 hr observation period is only 23 seconds, but the difference during the peak part of the observation period is approximately 100 seconds. The analysis period of oversaturated conditions must be long enough to include periods of unsaturated conditions at the beginning and at the end of the observation, but at the same time these two periods dilute the impact of the congestion period.

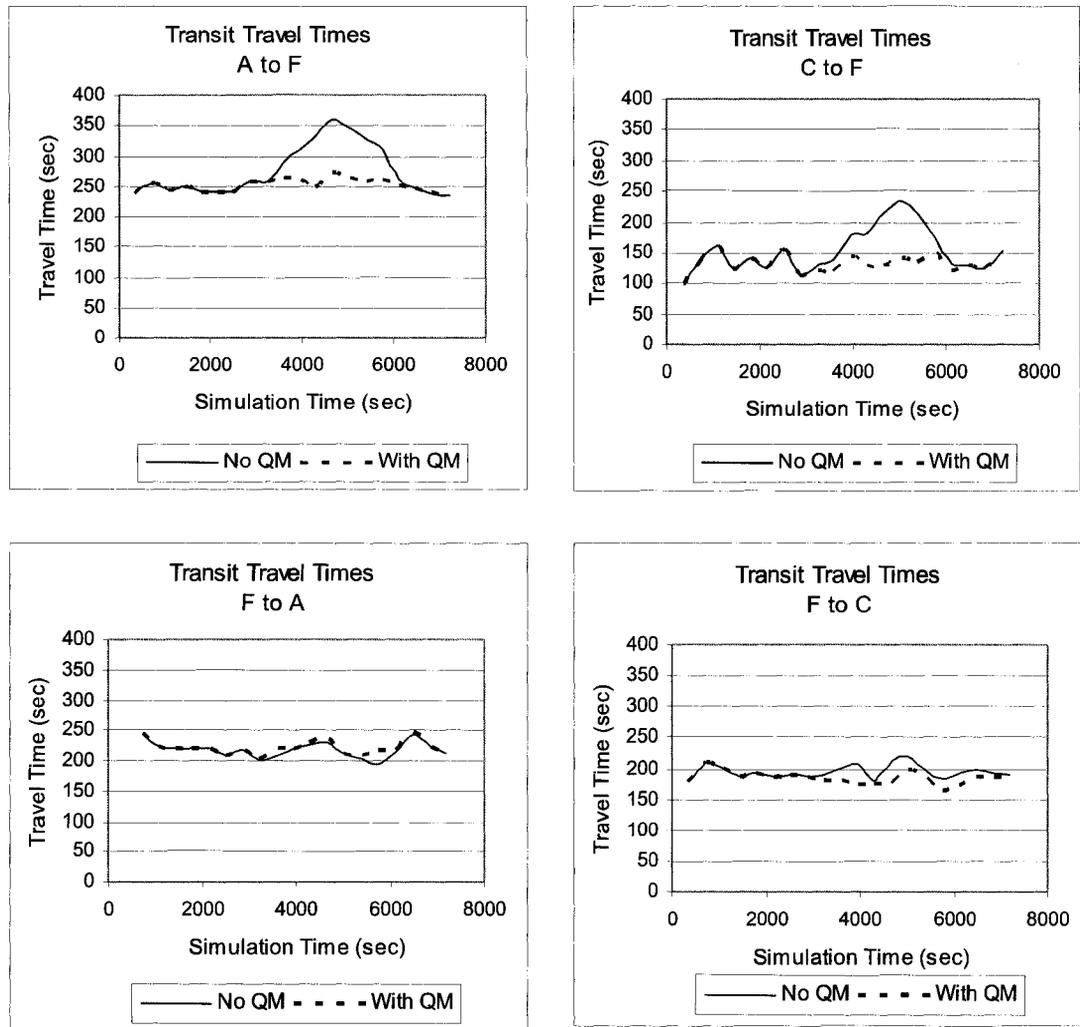


Figure 9.21: Comparison of Transit Travel Times During the Simulation Period

Transit vehicles normally operate by following time schedules. Time schedules define the time when transit vehicles should reach specific points along their routes.

The MoP used for scheduling purposes must include the variability of travel times as well. One approach to include both the mean travel time and its variability in a single quantity is to use percentile travel time. For example, the 85th percentile travel time would include 85% of all trips (Section 5.4.3).

The impact of queue management on transit travel time variability is explored by examining transit travel time distributions. Figure 9.22 depicts the travel time distribution for section A to F without queue management. The distribution is skewed to the right and the 85th percentile travel time is approximately 350 sec. With queue management there is much less skewness and the 85th percentile travel time is approximately 280 seconds (Figure 9.23). The travel time saving from the transit planning and operations perspective is 80 seconds.

Travel time A to F (sec)	No QM	With QM	Difference
Mean (bus)	270	246	24
85 th percentile (bus)	350	280	70

The real transit benefit of queue management calculated by using the 85th percentile is almost 3 times more than the benefits calculated by using the mean travel time savings which were 24 seconds (Table 9.8).

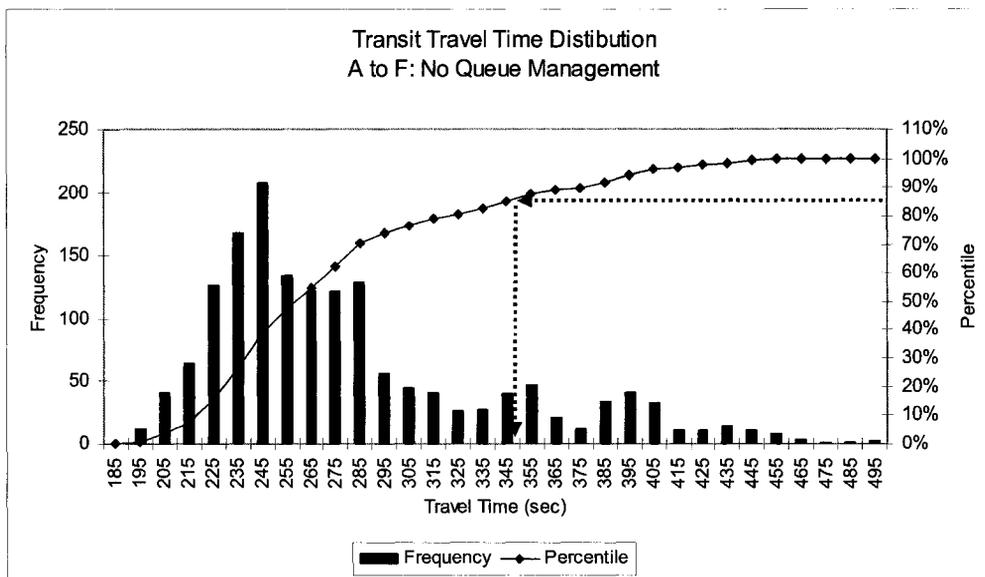


Figure 9.22: Transit Travel Time Distribution without Queue Management and the 85th Percentile Travel Time

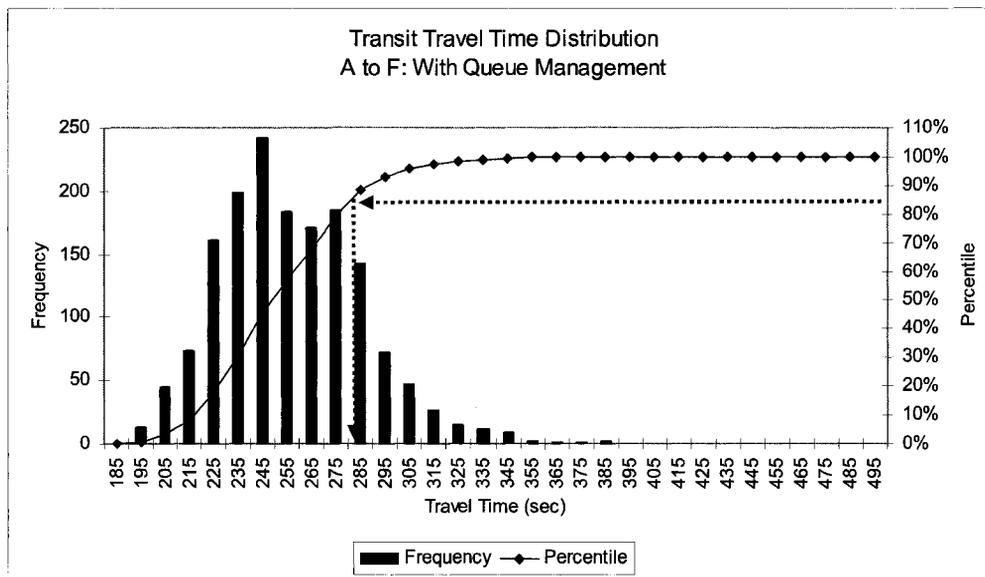


Figure 9.23: Transit Travel Time Distribution with Queue Management and the 85th Percentile Travel Time

9.4.3.3 Sensitivity to Different Levels of Congestion

To investigate the impact of queue management in various levels of congestion, traffic flow through the simulated bottleneck intersection in the corridor (Woodroffe Avenue and Baseline Road) was modified indirectly by changing the departure speed at the intersection (Step 6 in Section 9.4.1). Traffic flows were increased by approximately 5% to simulate lower levels of congestion and decreased by approximately 5% to simulate higher levels of congestion.

Table 9.9: Congestion Level Scenario Definitions

Congestion Level Scenario	NB Flow at Woodroffe and Baseline (veh/hr)	Index
Base Case Congestion	930	1
Decreased Congestion	970	1.43
Increased Congestion	880	0.95

The upper limit of the congestion level could not be increased any further without expanding the geographical scope of analysis area. As discussed earlier, the prerequisite for the analysis of oversaturated conditions is to keep the impacts of control changes within the scope of the simulated network. Queue formation is very sensitive to the level of congestion thus the range of traffic flows through the simulated bottleneck intersection that could be used in the analysis, is rather narrow.

Figure 9.24 depicts the change in the mean car and transit travel times as the level of congestion increases for the scenario without queue management. Both travel times increase, but the transit travel time increase is smaller because buses are impacted by

congestion along only 240 meters within the simulated corridor compared to cars that are impacted by congestion along the entire length of their travel (Figure 9.11).

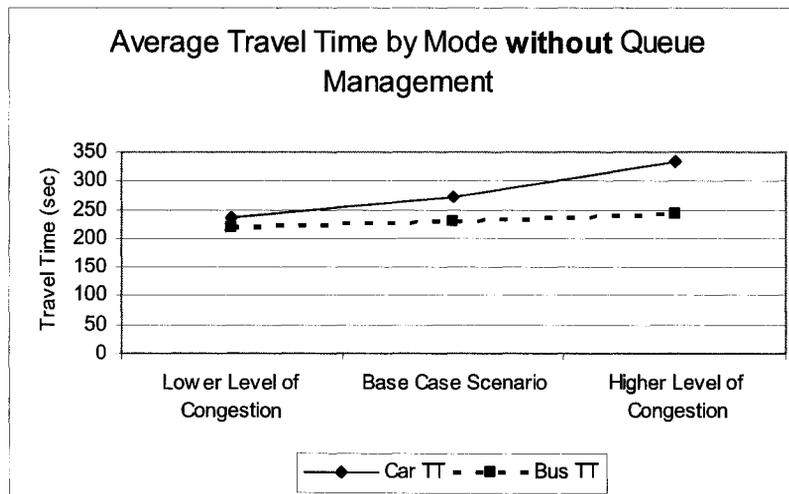


Figure 9.24: Impacts of the Level of Congestion on Car and Bus Travel Times without Queue Management

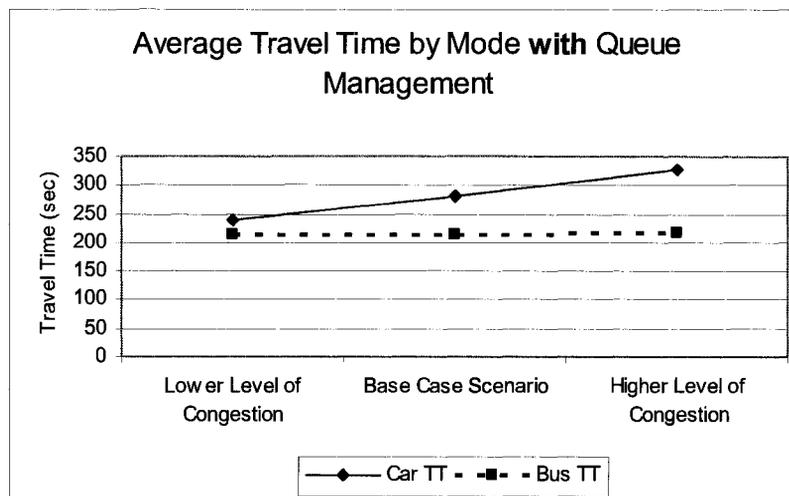


Figure 9.25: Impacts of the Level of Congestion on Car and Bus Travel Times with Queue Management

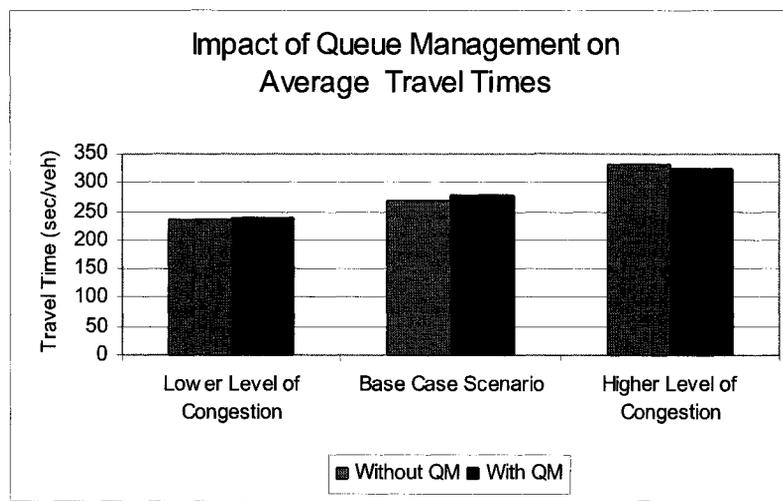


Figure 9.26: The Impacts of Queue Management on the Overall Mean Travel Time at Various Levels of Congestion

Figure 9.25 shows the change of car and transit mean travel times as the level of congestion increases for the scenario with queue management. There is an increase in car travel times as congestion increases and the change is very similar to the scenario without queue management (Figure 9.24). However, mean transit travel times remain constant within the simulated range of congestion levels.

Figure 9.26 and Figure 9.27 depict the impact of queue management on the overall effectiveness of the transportation system without differentiating between the two modes.

Although mean travel times calculated through simulation are not statistically significant (95% confidence level), they indicate that at lower levels of congestion, the mean travel time for all vehicles in the system are lower without queue management. However, this changes as the level of congestion increases (Figure 9.26).

In terms of total person travel time, the total time decreases with queue management as the level of congestion increases. However, from the statistical perspective, these changes are not significant (95% confidence level).

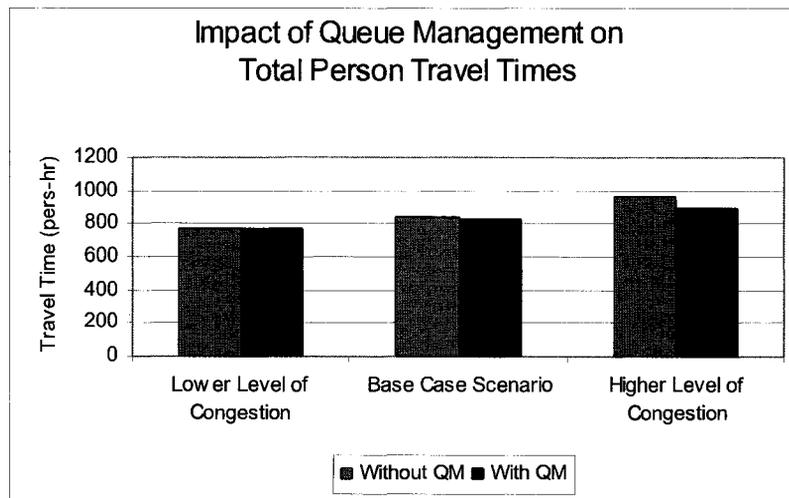


Figure 9.27: The Impacts of Queue Management on Total Person Delay at Various Levels of Congestion

9.4.3.4 Impact on Signal Controller Operation

In the previous section the performance of queue management algorithm was analysed and discussed from the transportation operations perspective. This section includes a brief analysis of the queue management performance from the signal control perspective.

The control objective of the queue management algorithm is to keep the number of vehicles within the detection zone (NVDZ) at a predefined target level by changing the green time at the upstream metering intersection. The detection zone and the target NVDZ is set so that during unsaturated conditions, the target NVDZ cannot and should not be achieved. During unsaturated conditions, the queue management algorithm

continuously tries to increase the green time at the metering intersection, however, the value of green time is limited by the programmed maximum green time. This ensures that the queue management component of the controller algorithm remains ineffective during unsaturated conditions. Therefore, there is no need for a special procedure that turns the congestion management algorithm on and off.

Figure 9.28 shows the variability of the NVDZ without queue management. The variability is relatively small since the detection zone is consistently filled up with vehicles. The green time is constant since there is no queue management algorithm.

Figure 9.29 shows the variability of the NVDZ and the green times adjustments with queue management. The NVDZ is oscillating around the target NVDZ and the green time varies between the minimum and maximum set green times.

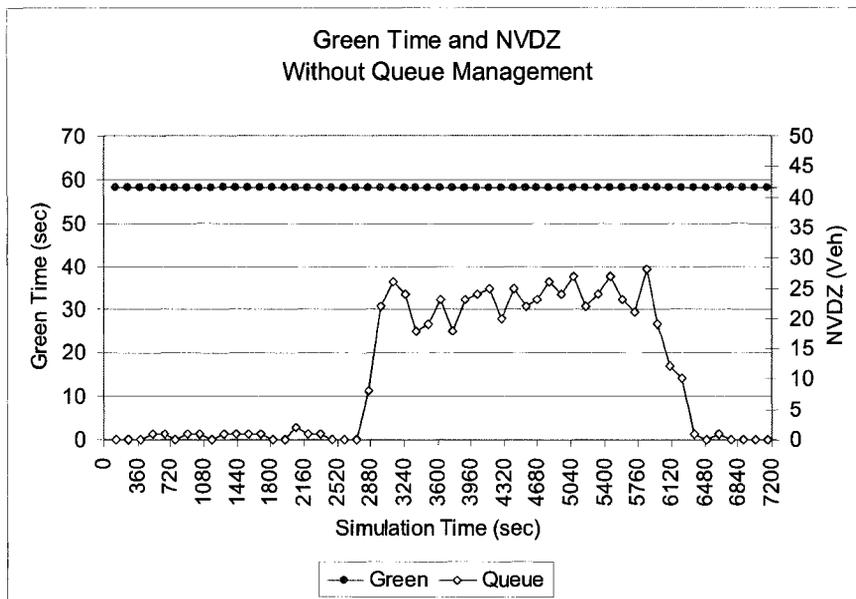


Figure 9.28: Green Time and NVDZ Variation during the Simulation Period without Queue Management

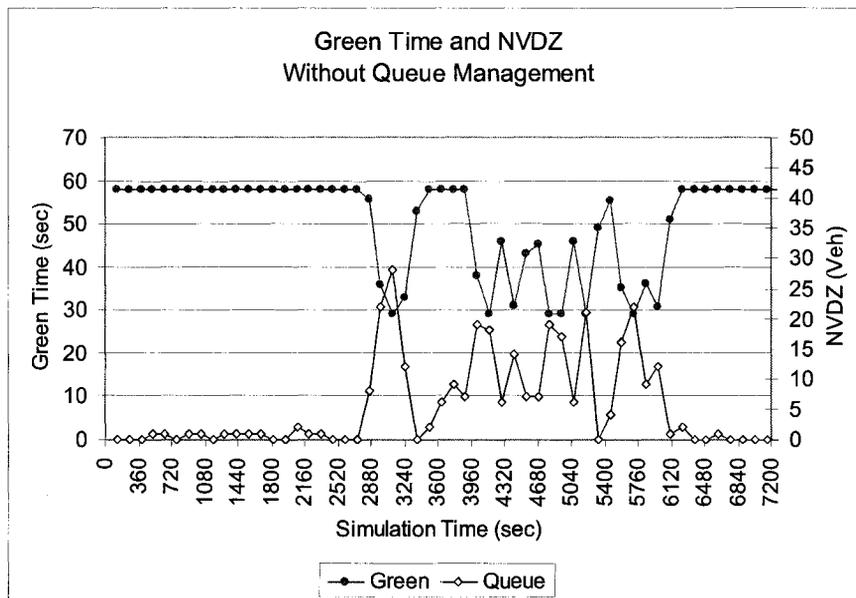


Figure 9.29: Green Time and NVDZ Variation during the Simulation Period with Queue Management

9.5 Summary

In this section the design and development of a queue management scheme upstream from the capacity bottleneck is provided. The identification and simulation of capacity bottlenecks along oversaturated corridors is demonstrated. The V/C calculation is useful but insufficient, empirical knowledge about operation in the wider area is also required. The geometric set-up and the queue management control algorithm is presented and applied on a simulated real-world example. Finally, travel time data were collected through simulation for pre-defined sections and they were compared for the operation with and without queue management.

Queue management at the upstream intersection (compared to the pre-signal) is likely to impact travel times on more travel segments. Aggregate measures of impacts on travel time are not too useful because they hide the trade-offs between travel times on various segments. The analysis of all impacted segments is required which significantly increases the complexity of evaluation.

The beneficial impact of bimodal queue management on transit travel times is evident, especially if the reduction of transit travel time variability is also included in the analysis. As expected, the impact on the overall vehicle travel times remains insignificant.

10. ESTIMATION OF THE NUMBER OF VEHICLES IN THE DETECTION ZONE

10.1 Introduction

The proposed queue management techniques are based on measurements of the number of vehicles approaching an intersection. This quantity is defined as the Number of Vehicles in the Detection Zone (NVDZ).

In this chapter, the development of a Fuzzy Logic Inference System (FIS) for the estimation of the NVDZ is described along with the Adaptive Neuro Fuzzy Inference System (ANFIS) based tuning process.

Finally the queue management algorithm using the NVDZ estimates is compared to the queue management algorithm based on vehicle counts developed in the previous chapter.

10.2 The Need for Estimation of NVDZ

In the queue management algorithm, the NVDZ serves as an indicator of the likely queue length that builds up at an intersection, but it does not correspond to any of the traditional definitions of queue length. The definition of a queue assumes stopped or slow moving vehicles. However, for queue management purposes developed in this thesis, it is sufficient to know only the number of vehicles that are approaching the intersection regardless of their speed.

The measure of *density*, used to describe the number of vehicles per unit distance, is another quantity which is similar to the NVDZ. The difference between *density* and NVDZ is more than the discrepancy of the units of measurement. *Density* is usually associated with uninterrupted flow environment where the objective is to describe average conditions at a longer roadway section. In the queue management application developed in this thesis, we are interested in the number of vehicles in a specific section with a limited length, usually a few hundred meters. The sections for NVDZ are selected to indicate the potential impacts of vehicular queues thus they do not represent the average density between two intersections.

One way to quantify the number of vehicles within the detection zone is by counting them as they cross specified entry and exit points and by taking the difference between the two numbers. The detectors at the entrance and exit to the detection zone are called check-in and check-out detectors respectively. Although theoretically accurate and applicable in a simulation environment, this approach has significant real-world implementation challenges.

Every vehicle detection technology has a certain error margin. Inductive loop detection is generally considered to be one of the most reliable technologies, but even its detection accuracy is less than 100%. Many traffic engineering applications can tolerate some detection inaccuracies but in an application where the number of vehicles is required between two detection points over or under counting is a significant problem. The reasons for detector over or undercounting include the following:

- on multilane roadways vehicles may pass between the detectors in the two adjacent lanes, or they may trigger both detectors as they change lanes, resulting in undercounting or over-counting;
- vehicles turning on and off the road within the detection zone are counted only at the exit or entry point;
- vehicles on two wheels can easily be counted at one location and missed at the other;
- when queues develop, vehicles following very closely may be missed especially if the detectors are longer (>3 m).

The identification of malfunctioning detectors is not always simple. If a detector is totally unoperational and provides no counts at all, identification is simple, but the identification of intermittent malfunctioning is problematic.

Consistent under or over-counting is a problem not only in real-world application but sometimes in a simulation environment as well. Even microscopic simulation programs such as VISSIM require special care to minimize and correct for over or undercounting detectors. The following factors were found to impact the operation of counting loops in VISSIM:

- too short detectors on multi-lane roadways may increase the chance of undercounting due to vehicle lane change;
- too long detectors on multi-lane roadways may increase the chance of over-counting due to vehicle lane changes;

- too long detectors may undercount in congested conditions because they increase the chance of being covered by two consecutive vehicles at the same time; and
- some transit vehicles are counted even if VISSIM has been set up to ignore them (possibly a software bug).

Through experimentation it has been found that detector lengths of 1 or 2 metres placed at locations with minimum lane changes provide the lowest counting error in VISSIM. In one example, the accumulated error was between 10 to 30 vehicles at the end of a four hour simulation period on a two lane roadway with approximately 1500 vehicles per hour. The 30 vehicles represent only a 0.5% cumulative counting error, however, if the capacity of the detection zone is 30 vehicles, the erroneous data may indicate that the detection zone is full when it is actually empty.

Unless the cumulative error can be corrected manually or by a special algorithm, any queue management technique that is based on the number of vehicles within a specific zone will gradually lose its functionality after a few hours in operation in either real-world or simulation environments. A more robust measurement of the NVDZ is required.

Papageorgiu et al. (2008) investigated the relationship between time-occupancy measurements, space-occupancy measurements, and link vehicle-counts for both uninterrupted and interrupted flow. They found that in the case of uninterrupted flow, both occupancy types are roughly proportional to traffic density. However, in case of signalized roadways, they found that the relationship was much more complex. Vigos (2008) builds on the results of Papageorgiu et al. and employs Kalman-Filter to produce

estimates of vehicle counts for signalized links based on three loop detectors. The link volume estimator is tested in a simulation environment and favorable results are reported.

10.3 Motivation for Fuzzy Logic Based NVDZ Estimation

Suppose the objective is to determine the number of vehicles within the detection zone (NVDZ) shown in Figure 10.1. Within the detection zone, traffic may be moving, may be stopped, or a combination of the two: part of it is moving and part of it is stopped.

Rectangles A, B, and C are vehicle detectors, (e.g. inductive loop detectors embedded in the pavement) which measure the percent of time within a short time interval while they are covered by a vehicle. This quantity is defined as the detector time occupancy.

The proposed NVDZ estimation process is based on the assumption that there is a relationship between the number of vehicles in the detection zone and detector occupancies at the three locations. To analyse this relationship data were collected through simulation. Detectors A, B, and C were used to collect occupancy information and the check-in/check-out detectors were counting the number of vehicles within the detection zone (Figure 10.1).

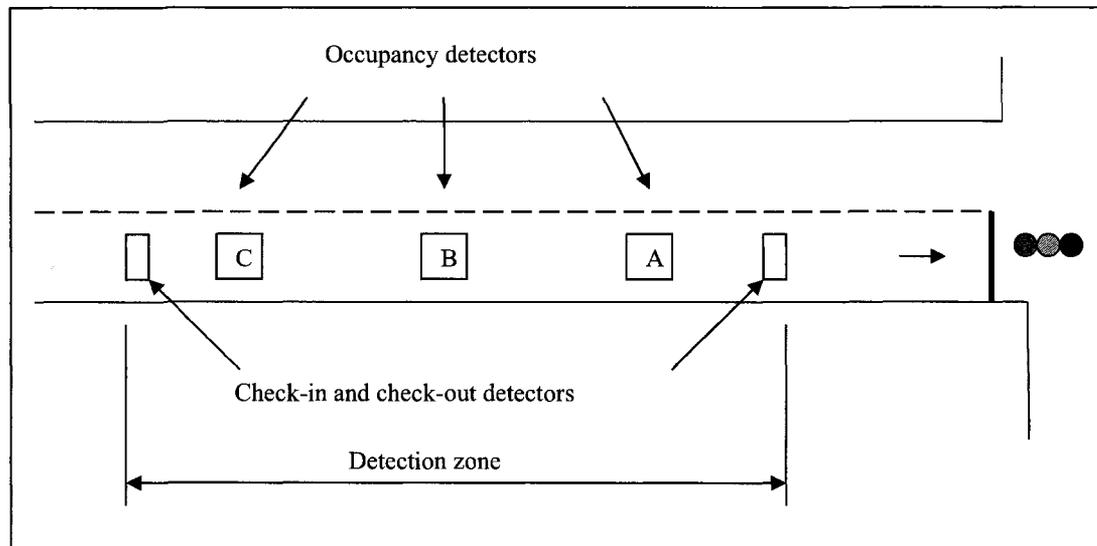


Figure 10.1: Detection Set-up for NVDZ Estimation

Data was collected every second for a four hour period (14400 data records). A signalised intersection was placed downstream from the detection zone and input volumes were increased and decreased every 900 seconds to generate different queue lengths within the detection zone. Sample records of the collected data set are shown in Table 10.1.

For example, at the 4827th second, detector A was covered for 96 % of the sampling interval (which was one second), detector B was covered for 16% and detector C was covered for only 7 %. In other words the vehicle above detector A was moving very slowly compared to vehicles above detector B and detector C. During the same second, the number of vehicles within the detection zone (NVDZ) was 14. At the 7421st second, occupancies increased at all three detectors to 99%, 99%, and 87% for the A, B, and C detectors respectively. The 99% actually means that the detector was covered during the

entire interval. Traffic was stopped or moved very slowly over all three detectors and the corresponding NVDZ was 43 vehicles.

Table 10.1: Sample Records of Detector Occupancies and Corresponding NVDZ

Seconds	Occupancy A	Occupancy B	Occupancy C	NVDZ
...				
4826	0.96	0.16	0.05	13
4827	0.96	0.16	0.07	14
4828	0.96	0.15	0.08	12
4829	0.96	0.14	0.08	12
4830	0.97	0.13	0.07	12
4831	0.97	0.12	0.06	14
...				
7419	0.99	0.99	0.86	42
7420	0.99	0.99	0.86	43
7421	0.99	0.99	0.87	43
7422	0.99	0.99	0.88	44
...				

The output data for the four hour analysis period was plotted for a preliminary analysis.

First, detector occupancy data were plotted against the NVDZ for each occupancy detector individually (Figure 10.2, Figure 10.3, and Figure 10.4). Second, detector occupancies at all three locations were plotted against NVDZ (Figure 10.5).

Based on the individual detector plots one can make some conclusions about the range of the NVDZ for various detector occupancies. For example, when the occupancy of detector A is 0.2, the NVDZ is between 3 and 17 vehicles and the range of variation is approximately 14 vehicles (Figure 10.2). When occupancy of detector A is close to 1.0, the NVDZ is between 7 and 45 vehicles and the range of variation is much larger,

approximately 38 vehicles. Thus detector A can provide better NVDZ estimates when the queue length is less than 25 vehicles.

Similarly, observation can be made for detector B (Figure 10.3) and detector C (Figure 10.4). When detector B occupancy is between 0.4 and 0.8, the NVDZ is between 20 and 35 vehicles. When detector C occupancy is above 0.6, the NVDZ is between 35 and 45 vehicles

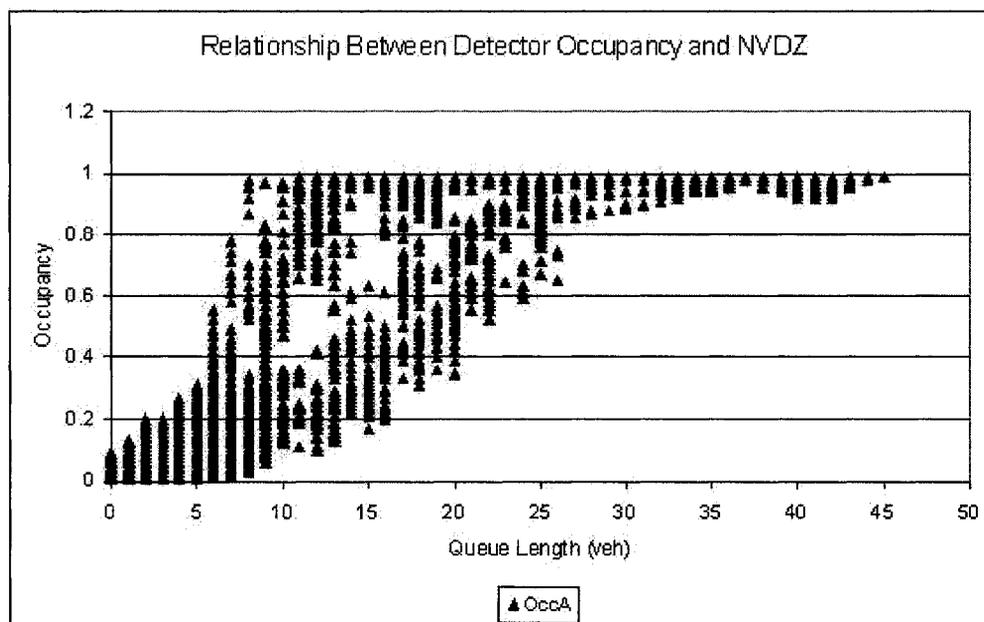


Figure 10.2: The Relationship between Detector A Occupancy and NVDZ

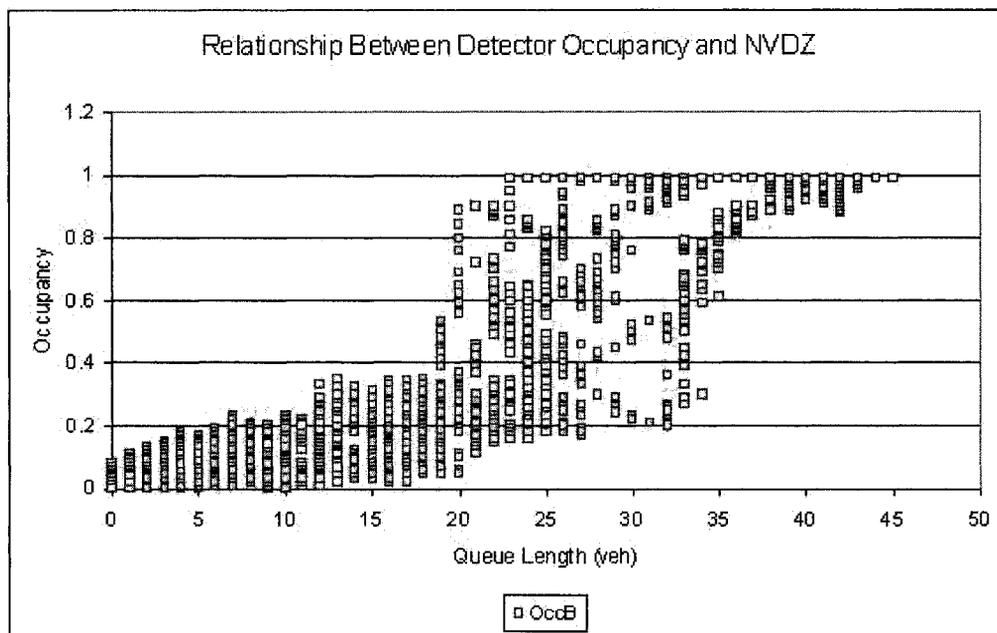


Figure 10.3: The Relationship between Detector B Occupancy and NVDZ

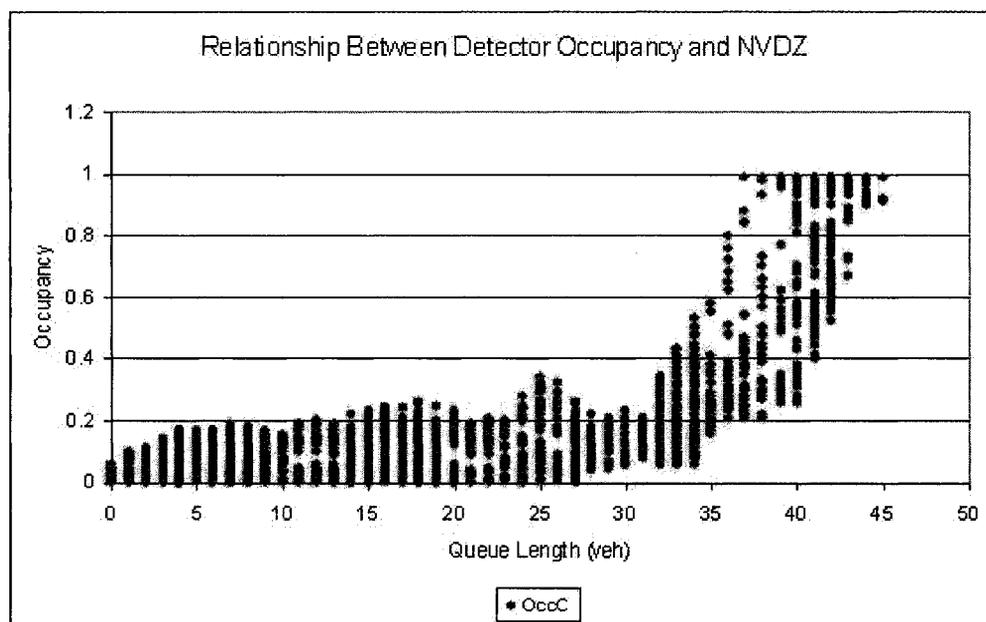


Figure 10.4: The Relationship between Detector C Occupancy and NVDZ Length

By utilising all three detector information individually, it is possible to draw some limited conclusions about the value of the NVDZ (Table 9.2). However, the estimation of the NVDZ based on individual detector occupancies is problematic. First, the range of the NVDZ for each detector is rather large. Second, one does not know which detector information to use when the NVDZ outcomes are different. Third, the occupancies in Table 10.2 do not cover the entire range of possible occupancy values.

Table 10.2: NVDZ Estimation Based on Individual Detector Occupancies

Detector	Occupancy (%)	NVDZ (veh)
A	<0.2	0-15
A	0.4-0.8	5-25
B	0.4-0.8	25-35
C	0.6-1.0	35-45

By combining the occupancy information of all three detectors it is expected that one would be in a better position to determine the NVDZ more accurately. However, visualizing the three inputs (occupancies on detectors A, B, and C) and one output (NVDZ) in a way that would make intuitive decisions possible is difficult. Figure 10.5 is an attempt to show all four variables in one graph. Certain patterns are obvious and to describe those patterns a structured and systematic analysis and decision process is required.

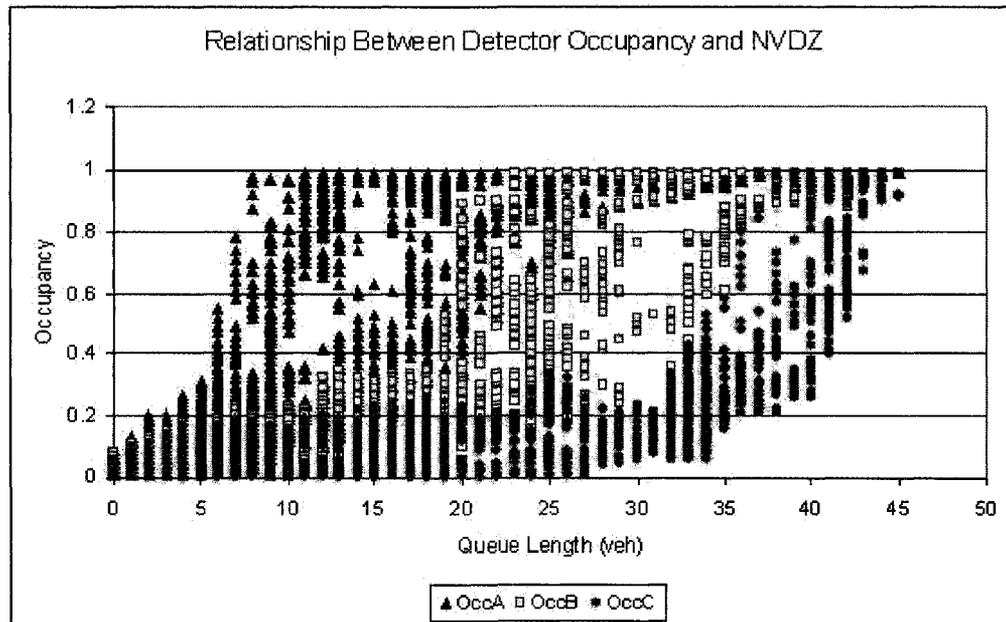


Figure 10.5: The Relationship between all three Detector Occupancies and NVDZ

The relationship between detector occupancies and the NVDZ (Figure 10.5) can be described by fuzzy rules. Assume that OccA, OccB, and OccC represent the occupancies of detectors A, B, and C within the sampling time interval. It is obvious that higher occupancy on all three detectors mean bigger NVDZ. This is described by the following fuzzy rule:

Rule 1:

If OccA is High AND OccB is High AND OccC is High THEN NVDZ is Very Big

The other extreme case is when occupancies are low or zero and there is a short or no queue. This can be described by the following fuzzy rule:

Rule 2:

If OccA is Low AND OccB is Low AND OccC is Low THEN NVDZ is Very Small

Between these two extremes, there are many possible NVDZ values which require the definition of a number of rules. Some of these rules are intuitively simple, such as the following two:

Rule 3:

If OccA is High AND OccB is Low AND OccC is Low THEN NVDZ is Small

Rule 4:

If OccA is High AND OccB is High AND OccC is Low THEN NVDZ is Medium

However, the queue formation and dissipation dynamics creates situation which are difficult to define intuitively. For example, one would need significant understanding of a queue formation to determine whether or not the NVDZ for Rule 5 should be bigger than the NVDZ for Rule 6.

Rule 5:

If OccA is Medium AND OccB is Medium AND OccC is High THEN NVDZ is “must be determined”

Rule 6:

If OccA is High AND OccB is Medium AND OccC is Medium THEN NVDZ is “must be determined”

With three occupancy detectors and three membership functions for each detector (e.g. Low, Medium, and High) 27 rules could be formed. Determining the corresponding NVDZ for each rule would require significant knowledge of the queue formation and dissipation dynamics, lots of experimentation, and/or analysis of the data in Figure 10.5.

This approach is not pursued. Instead, a different method for developing a Fuzzy Inference System (FIS) is proposed.

10.4 Fuzzy Inference System Development for NVDZ Estimation

In the previous section, the existence of a relationship between detector occupancies and the corresponding number of vehicles in the detection zone (NVDZ) has been established. Furthermore, the potential of describing this relationship by using fuzzy rules has been suggested but the intuitive approach of developing fuzzy rules and tuning membership functions has significant challenges. Fortunately, it is possible to collect data which describes the relationship between the inputs (detector occupancies) and output values (NVDZ) by simulation. Thus, the idea is to use a technique which can learn from the available data and which can be used to develop a Fuzzy Inference System (FIS). A computational framework, developed by Jang (1993), is proposed for this purpose.

10.4.1 Network Based Learning Systems

The application of network based or neuro based learning for various transportation problems have been suggested by a number of researchers. Teodorović et al. (1998) described a number of potential applications of fuzzy logic and neural networks in transportation problems. Park (2002) described the application of a hybrid network based-fuzzy system for short-term freeway traffic volume forecasting. Přibyl et al. (2003) applied adaptive neuro fuzzy inference system in the analysis of travel behaviour.

Andrade et al. (2006) developed an adaptive network based fuzzy logic transportation choice model.

The idea behind the neuro-network is to create a computing structure that is similar to the functioning of the human brain. The brain is a collection of about 10 to 100 billion interconnected neurons. Each neuron is a cell that uses biochemical reactions to receive, process, and transmit information. Neurons send information to other neurons in the form of a positive or negative charge. The strengths of all the received charges are added together through the processes of spatial and temporal integration. Spatial integration occurs when several weak signals are converted into a single large one, while temporal integration converts a rapid series of weak pulses from one source into one large signal. The aggregate input is then passed to other neurons. If the aggregate input is greater than a specific threshold value, an output signal is transmitted to other neurons (Fraser, 1998; Neuron, 2008).

Due to technical limitations, the artificial neuro-network is much simpler than the biological neuro-network. In an artificial neuro-network, neurons or nodes are organized into distinct layers. There are three layer categories, input, hidden, and output layers. Neurons in the input layer receive signals from outside sources, neurons in output layers transmit signals outside the network. Neurons in the hidden layer are not “visible” from the outside. Neurons in one layer are connected to all neurons in the next layer (Figure 10.6). Information is flowing only in one direction hence the name Feed-Forward Neural Network Model.

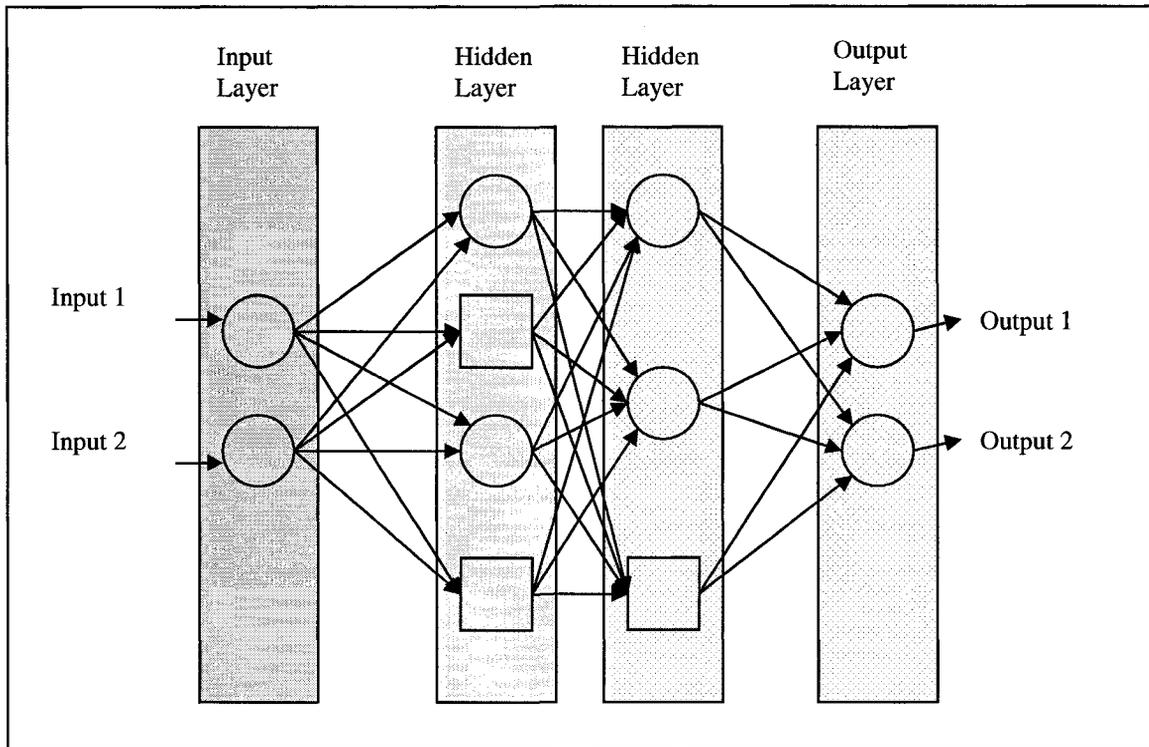


Figure 10.6: Architecture of a Simple Artificial Neuro-Network

Every node has three functions: accumulation, multiplication, and activation (Figure 10.7). The accumulation function collects all input values (whether they are external inputs or inputs from previous layers). The multiplication function calculates the weighted average by multiplying the input values with the weights of the corresponding connections. The activation function modifies the signal(s).

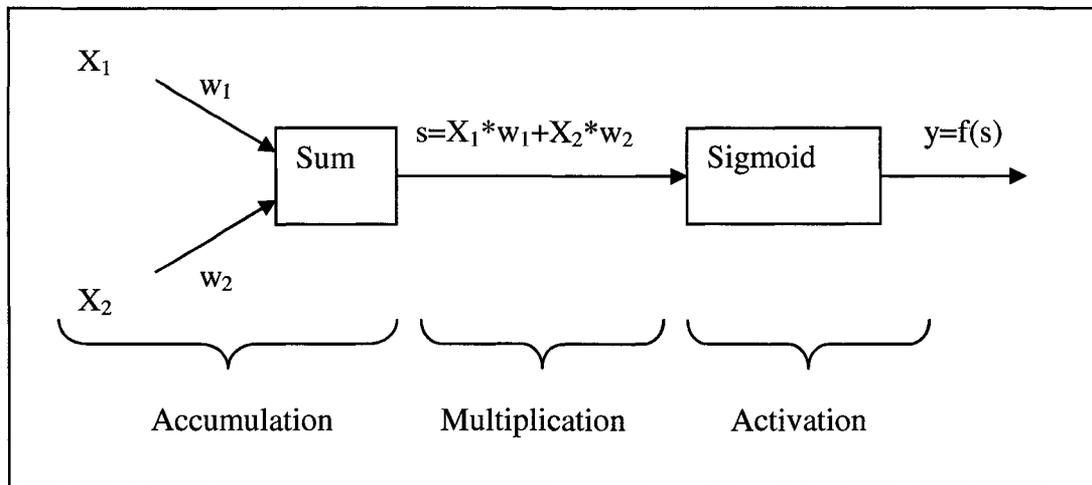


Figure 10.7: The Structure and Functions of an Artificial Neuron/Node

The most commonly used activation function is the sigmoid function (logistic curve) described by Equation 10.1 and depicted in Figure 10.8. However, other activation or node functions can be used and the only limitation is that they must be piecewise differentiable.

$$S = f(s) = \frac{1}{1 + e^{-s}} \quad 10.1$$

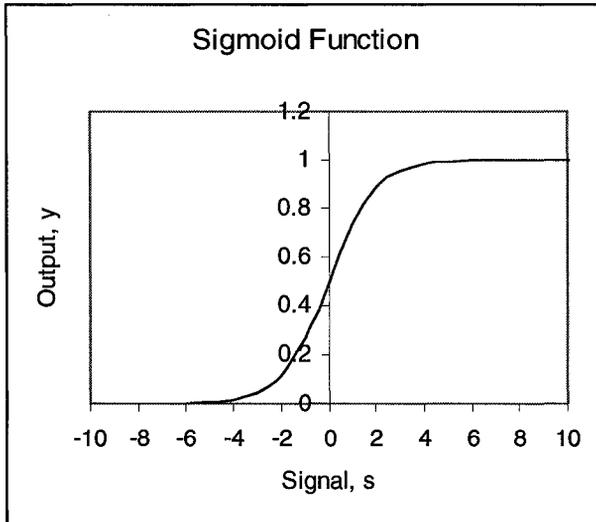


Figure 10.8: Sigmoid Function (Logistic Curve)

There are at least two reasons for using the sigmoid function for signal activation. First, the sigmoid function enables the reception of both very weak and very strong signals (Teodorović et al., 1998). When the input signal is weak, the curve slope is large. When the signal is very strong, the curve slope is small. Second, the sigmoid function simplifies mathematical data manipulation since the derivative of the sigmoid function can be expressed in terms of the function itself (10.2).

$$\begin{aligned} \frac{d}{ds} \left(\frac{1}{1+e^{-s}} \right) &= \frac{e^{-s}}{(1+e^{-s})^2} = \frac{1+e^{-s}-1}{(1+e^{-s})^2} = \frac{1}{1+e^{-s}} - \frac{1}{(1+e^{-s})^2} = \\ &= \frac{1}{1+e^{-s}} \left(1 - \frac{1}{1+e^{-s}} \right) \end{aligned} \quad 10.2$$

The “knowledge” of the network is determined by the connections’ weights which are used to multiply individual signals. The objective of neuro-network training is to adjust these weights. The neuro network is trained by feeding the network with a set of input

numbers and the network produces output numbers based on the initial weights. The output is then compared to the corresponding training data and the error is calculated.

Since the weights are initially set randomly, the error will likely be significant. To reduce the error, the weights of some of the connections are adjusted and the output is calculated again. This is repeated for each data point in the training data set. Then the whole process is repeated until the error becomes sufficiently small. To determine which connections' weight needs to be adjusted and by how much to minimize the output error, the back-propagation algorithm can be used.

The main steps of determining the network output and output error are explained through a graphical example and then generalized by formulas.

Figure 10.9 shows a simple network with three layers and two nodes in each layer. In general, neural networks can have any number of layers and any number of nodes within the layers. The principles, described below apply to networks of any size.

In this simple example, and X_i represents the inputs, and Y_i and O_i represent the outputs of the hidden and output layers respectively. The weights between nodes i and j are marked by w_{ij} and the ^H and ^O superscripts indicate that they belong to the hidden or the output layer.

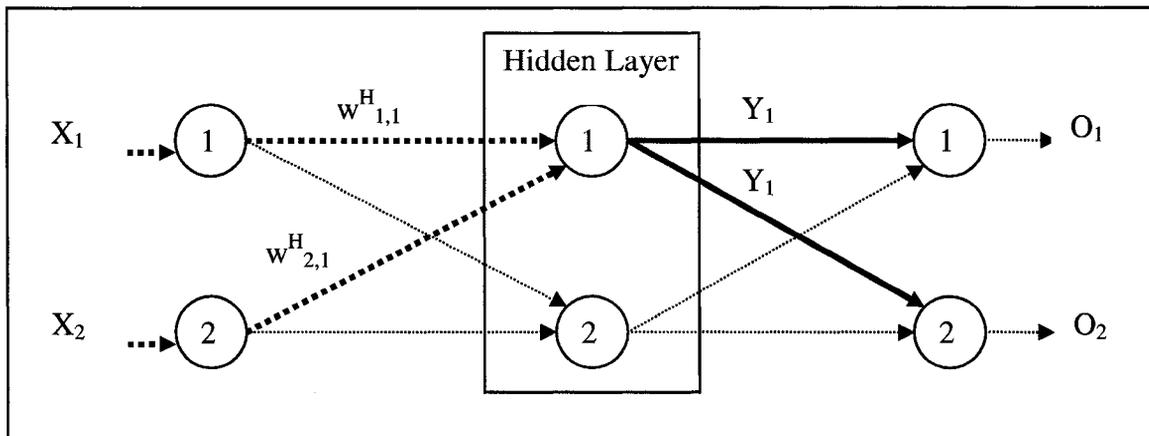


Figure 10.9: Output Calculation for Node 1 in the Hidden Layer

The output of Node 1 in the hidden layer is determined by applying the sigmoid function to the sum of products of the input values and their corresponding weights (10.3).

$$Y_1 = S\{X_1 \times w_{1,1}^H + X_2 \times w_{2,1}^H\} = S\left\{\sum_{l=1}^2 X_l \times w_{l,1}^H\right\} \quad 10.3$$

S - represents the sigmoid function defined in Equation 10.1

l – index of node in the input layer

The calculation of the output of Node 2 in the hidden layer is presented graphically in Figure 10.10 and symbolically by Equation 10.4.

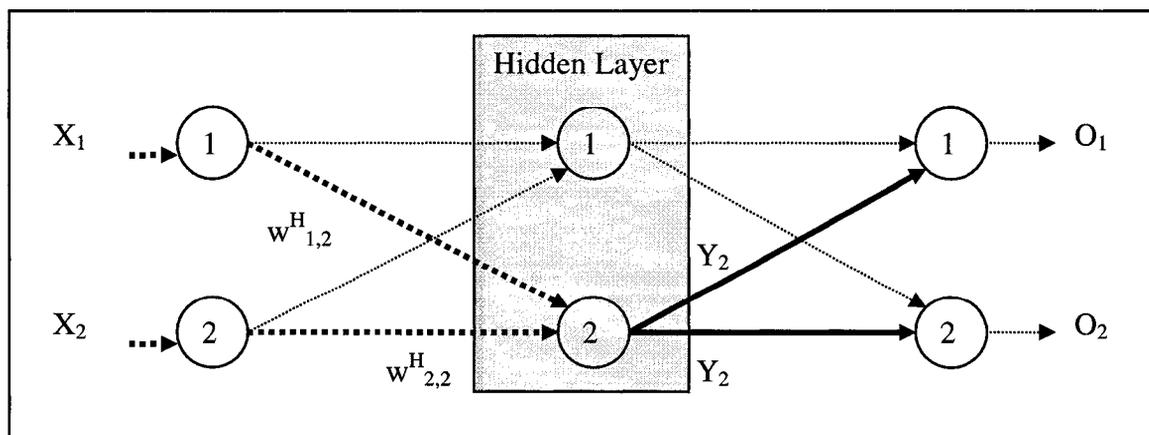


Figure 10.10: Output Calculation for Node 2 in the Hidden Layer

$$Y_2 = S\{X_1 \times w_{1,2}^H + X_2 \times w_{2,2}^H\} = S\left\{\sum_{l=1}^2 X_l \times w_{l,2}^H\right\} \quad 10.4$$

The process of determining Node outputs for the output layer is very similar. However, the inputs are now the outputs of the previous layer. Figure 10.11 and Figure 10.12 show the graphical presentation and Equation 10.5 and Equation 10.6 show the symbolic formulation.

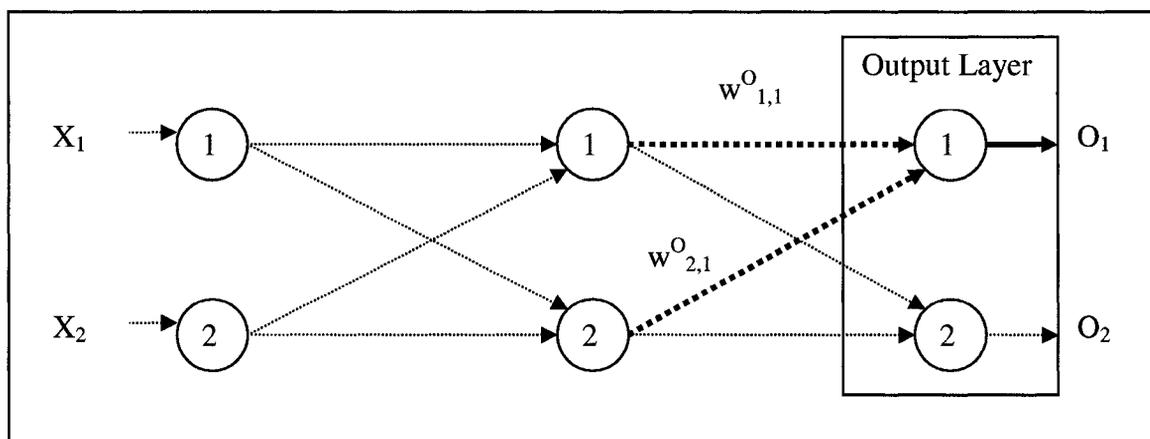


Figure 10.11: Output Calculation for Node 1 in the Output Layer

$$O_1 = S\{Y_1 \times w_{1,1}^O + Y_2 \times w_{2,1}^O\} = S\left\{\sum_{m=1}^2 Y_m \times w_{m,1}^O\right\} \quad 10.5$$

m – index of node in the hidden layer

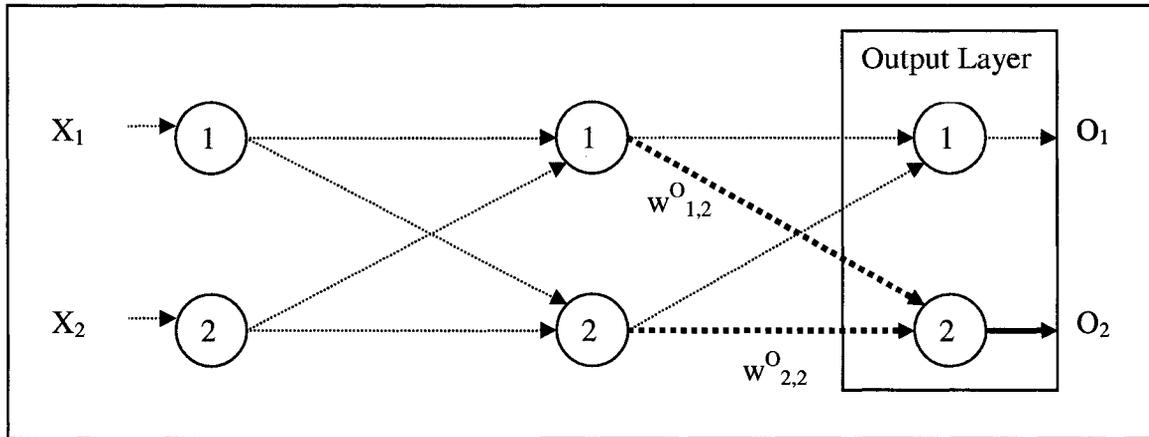


Figure 10.12: Output Calculation for Node 2 in the Output Layer

$$O_2 = S\{Y_1 \times w_{1,2}^O + Y_2 \times w_{2,2}^O\} = S\left\{\sum_{m=1}^2 Y_m \times w_{m,2}^O\right\} \quad 10.6$$

Equations 10.5 and 10.6 representing the outputs of the *output layer* contain the outputs of the *hidden layer*. By inserting 10.3 into 10.5, equation 10.7 is derived where the output O_1 is expressed as a function of network input values X_i and the weights w^O , w^H .

$$O_1 = S\left\{\sum_{m=1}^2 S\left\{\sum_{l=1}^2 X_l \times w_{l,m}^H\right\} \times w_{m,1}^O\right\} \quad 10.7$$

Similarly, by inserting 10.4 into 10.6, Equation 10.8 is derived which expresses the output O_2 as a function of network inputs and weights without the need to calculate the interim outputs in the hidden layer.

$$O_2 = S\left\{\sum_{m=1}^2 S\left\{\sum_{l=1}^2 X_l \times w_{l,m}^H\right\} \times w_{m,2}^O\right\} \quad 10.8$$

At this point it is possible to determine the error of the network calculation. The error is defined as the squared difference between the calculated network output and the measured output (10.9).

$$E = \sum_{n=1}^2 (D_n - O_n)^2 \quad 10.9$$

D_n – desired or measured output for data set or record n

By substituting O_n with 10.7 and 10.8, Equation 10.10 is derived which represents error as a function of the measured output, network inputs, and weights.

$$E = \sum_{n=1}^2 \left(D_n - S \left\{ \sum_{m=1}^2 S \left\{ \sum_{l=1}^2 X_l \times w_{l,m}^H \right\} \times w_{m,n}^O \right\} \right)^2 \quad 10.10$$

Equation 10.10 represents a multidimensional surface with parameters $w_{l,m}^H$ and $w_{m,n}^O$ which represent the weights of the hidden and output nodes for each link in the network. To minimize the error the weights have to be adjusted by moving to the lowest point on the error surface. To find the lowest point, the gradient of the error function has to be calculated. The gradient is defined by the partial derivative of the error function with respect to each variable.

The partial derivative in respect to the output layer weights is adapted from Crochat et al. (2003) and presented in Equations 10.11, 10.12, and 10.13.

$$\begin{aligned} \frac{\partial E}{\partial w_{m,n}^O} &= \frac{\partial}{\partial w_{m,n}^O} \times \sum_{n=1}^2 (D_n - O_n)^2 = -2 \times (D_n - O_n) \times ((1 - S\{W^O\}) \times S\{W^O\}) \times Y_m \\ &= \delta_n^O \times Y_m \end{aligned} \quad 10.11$$

where W^O is expressed by 10.12

$$W^O = \sum_{k=1}^2 w_{m,n}^O \quad 10.12$$

and δ_n^O is expressed by 10.13.

$$\delta_n^O = -2 \times (D_n - O_n) \times ((1 - S\{W^O\}) \times S\{W^O\}) \quad 10.13$$

Finally, the new weights are calculated by using Equation 10.14 where η is a new parameter indicating the step size of adjustment.

$$w_{m,n}'^O = w_{m,n}^O + \eta \times \frac{\partial E}{\partial w_{m,n}^O} \quad 10.14$$

η – step size parameter

Similarly, the partial derivatives in respect to hidden layers can be calculated. The final formula is presented in 10.15.

$$\frac{\partial E}{\partial w_{l,m}^H} = \frac{\partial}{\partial w_{l,m}^H} \times \sum_{n=1}^2 (D_n - O_n)^2 = ((1 - S\{W^H\}) \times S\{W^H\}) \times \sum_{n=1}^2 \delta_n^O \times w_{m,n}^O \times X_l \quad 10.15$$

where

$$W^H = \sum_{k=1}^2 w_{l,n}^H \quad 10.16$$

Finally the new weights for the hidden nodes are calculated by 10.17.

$$w_{l,m}'^H = w_{l,m}^O + \eta \times \frac{\partial E}{\partial w_{l,m}^H} \quad 10.17$$

The final equations for calculating the new weights for the output and hidden layers are based on the previous weight and the gradient of the error multiplied by a step size (10.14 and 10.17).

The gradient method described above is usually slow and likely to find only a local optimum. Therefore, Jang (1993) developed a hybrid learning rule which combines the gradient method and the least square estimate to identify parameters and is more likely to

find the global optimum. A detailed description of the method is provided by Jang (1993).

10.4.2 Adaptive Neuro-Fuzzy Inference System (ANFIS)

The structure of the neuro-network described in the previous section is functionally very similar to the structure of a Fuzzy Inference System (FIS). Jang (1993) applied the network learning algorithm to FIS and named the resulting architecture Adaptive Network Based Fuzzy Inference System (ANFIS). The same concept is often referred to with a slightly different name: Adaptive Neuro-Fuzzy Inference System (MathWorks, 2001).

ANFIS combines fuzzy logic and artificial neural networks. Fuzzy logic is used to model non-linear systems by a Fuzzy Inference System containing fuzzy membership functions, rule base, and a reasoning/inference mechanism (Section 6.3). The neural network is used to tune input and output membership function parameters.

The ANFIS method, implemented in MATLAB was used in this research. The FIS structure, including the number and shape of membership functions is set manually but the membership function parameters are adjusted by ANFIS through a learning process.

In MATLAB's ANFIS, the computation of the membership function parameters can be carried out by either the back-propagation algorithm or the hybrid algorithm based on a combination of back-propagation and a least squares estimator (MathWorks, 2001). For the purpose of this research, the hybrid algorithm was selected.

ANFIS is used to determine the number of vehicles within a detection zone based on occupancy detectors located within this zone. ANFIS uses input data (detector occupancies) and output data (NVDZ) to tune the parameters of membership functions.

The ANFIS technique, adapted in MATLAB, has the potential for using two data sets, one for training and the other for checking. The training data set is used to tune membership function parameters and the checking data set is used at the same time to assess the performance of the model and to prevent over-fitting.

During the learning process, ANFIS calculates the error defined as the sum of the squared differences between the calculated and measured value of NVDZ for both the training and checking data sets. During this process, the error for the training data set always decreases or remains constant. The error for the checking data, however, may start to increase at a certain point if the model loses its capability to generalize and starts to over-fit the data. If the checking data set is used, ANFIS reports membership function parameters that were generated before the checking data started to increase.

10.4.3 The Application of ANFIS for Parameter Tuning

The process of generating and tuning a FIS by using ANFIS consists of the following two main activities:

- Data collection and
- FIS generation.

Data Collection

Data collection consists of gathering sufficient *relevant* data describing the relationship between detector occupancies and the corresponding NVDZ values. The required data include detector occupancies at all three detection locations and vehicle counts at the check-in/check-out detectors. The *relevancy* of the collected data means, that it must cover a *range* of operating conditions, otherwise ANFIS cannot be trained for different conditions.

The *range* of operating conditions means collecting records (data points) with different combinations of occupancies at the three detection locations. This was partially achieved by periodically changing the input volumes so that queues would generate and dissipate within the detection zone a number of times during the simulation period.

A number of input volume scenarios were used during the evaluation process (Appendix T). This was required because different geometric set-ups and signal timings generate different queuing dynamics. To make sure that queue formation and dissipation occurs within the detection zone, input volumes were adjusted accordingly for every scenario tested.

For each scenario to be tested, a data base with 14400 records was created by collecting occupancy and NVDZ values every second during a VISSIM simulation run. The data base was then used as training data in ANFIS.

FIS Generation

The FIS was generated by using MATLAB's ANFIS tool. Figure 10.13 shows the structure of the Fuzzy Inference System which is similar to the structure of neuro network with five layers.

The Inputs (first layer) are the occupancy values of the three detectors (OccA, OccB, and OccC). Each input has three membership functions (Short, Medium, and Long) (second layer), and the complete Rule base has 27 rules (third layer). Each rule has singleton type Output membership function (fourth layer). The final Output is generated by defuzzification using the weighted average defuzzification method (fifth layer).

The structure of the FIS was constrained to allow only three membership functions (Short, Medium, and Long) for the occupancy input, and all input the membership functions were specified to have a trapezoid shape.

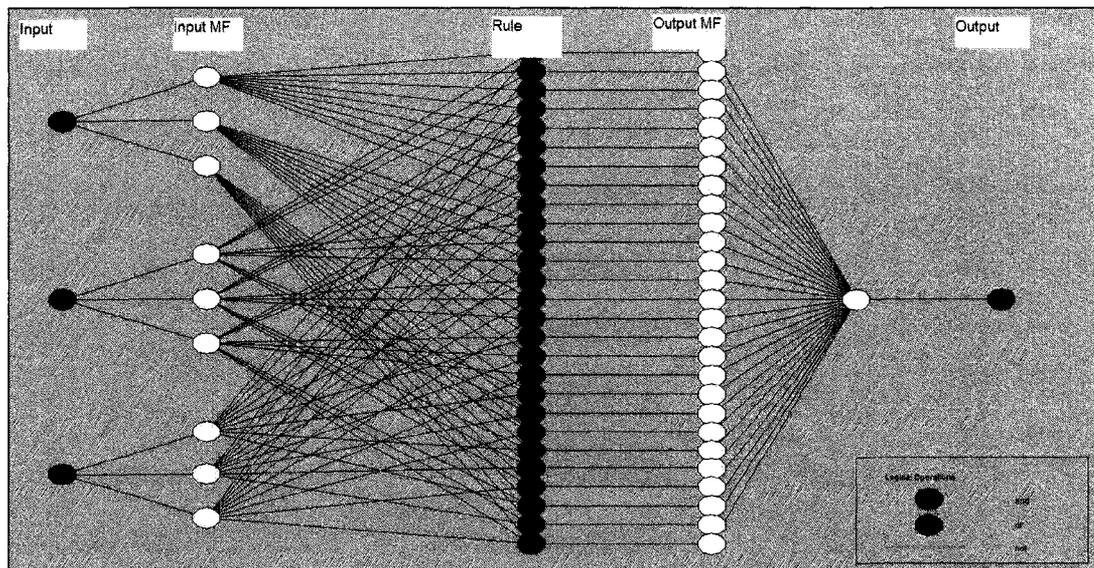


Figure 10.13: The FIS Model Structure Defined in ANFIS

The number of membership functions was restricted to three because having more than three membership functions significantly increases the number of rules but provides insignificant improvements in terms of reducing the error. For example, four membership functions for each input occupancy would require the definition of 48 rules ($3 \times 4 \times 4$) but the approximate error reduction was checked and was less than 2%.

The shapes of the membership functions for the input values were specified to have only a trapezoid form. The main difficulty of using other membership functions such as the Gaussian or generalized bell-shaped functions was mainly computational. The next step in the NVDZ estimation development process is testing and application (Figure 10.15) which is done in a VISSIM/VisVap environment. However, VisVap is a relatively simple programming language without the technical capabilities to describe Gaussian or generalized bell-shape functions. The output membership function was specified as a singleton value mostly to simplify later implementation in VISSIM/VisVap.

The error which is defined as the sum of the squared differences between the measured and calculated NVDZ values is shown in Figure 10.14 as the ANFIS is searching for the best FIS parameters. There is a sharp decrease of error during the first ten epochs and the error levels off after 50 epochs. (Epoch describes one sweep in ANFIS).

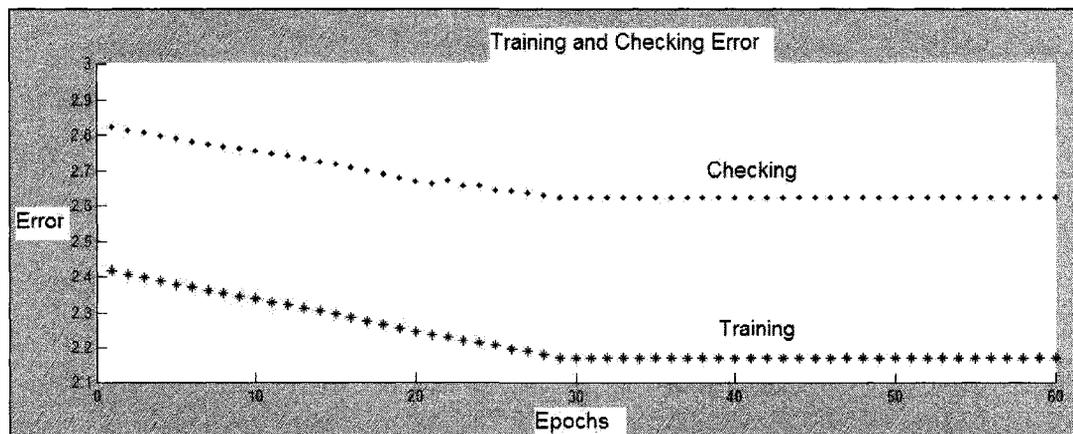


Figure 10.14: The Error for the Training and Checking Data

During the learning process, ANFIS modifies input and output membership function parameters (which are identical to the weights in neural networks). The final membership function parameters generated for the sample data set are shown in Table 10.3.

Table 10.3: Membership Function Parameters

Membership function	Parameters of the trapezoid			
	a	b	c	d
OccA_S	-99	-95	0.20	0.39
OccA_M	0.14	0.39	0.70	0.99
OccA_L	0.61	0.88	95	99
OccA_S	-99	-95	0.08	0.40
OccA_M	0	0.24	0.66	0.88
OccA_L	0.65	0.86	95	99
OccA_S	-99	-95	0.13	0.37
OccA_M	0	0.26	0.67	0.96
OccA_L	0.64	0.86	95	99

The rule base generated by ANFIS has 27 rules for the sample data set which are summarized in Table 10.4. The rule description is provided in the columns next to the Rule#, and the outcome of the rule is in the following column. For example for Rule 6 the

SML description means that the occupancy on detector A is Small, the occupancy on detector B is Medium, and the occupancy on detector C is large and the corresponding NVDZ is 15 vehicles.

Table 10.4: Rule Base Definition and the Corresponding Output Values

Rule#	Descr.	Output	Rule#	Descr.	Output	Rule#	Descr.	Output
1	SSS	1.6	10	MSS	4.6	19	LSS	10.5
2	SSM	6	11	MSM	21.2	20	LSM	29.3
3	SSL	20	12	MSL	25	21	LSL	35
4	SMS	0.4	13	MMS	12.8	22	LMS	20
5	SMM	18.9	14	MMM	17.7	23	LMM	30.4
6	SML	25	15	MML	22	24	LML	50.9
7	SLS	0	16	MLS	20	25	LLS	23.3
8	SLM	0	17	MLM	30.9	26	LLM	35.6
9	SLL	0	18	MLL	41.4	27	LLL	42.9

10.5 Evaluation of Fuzzy Logic Based NVDZ Estimation

The objective of evaluation is to quantify the accuracy and sensitivity of the Fuzzy Logic based NVDZ estimation technique. Accuracy describes the difference between the measured and estimated value of the NVDZ. Sensitivity describes the estimation techniques' performance in different geometric and operational conditions.

The evaluation process is described in Figure 10.15 and it is divided into three parts:

- The first part includes Step 1 (Data collection) and Step 2 (FIS development) which have already been discussed (Section 10.4.3).
- The second part includes Step 3, the transfer of the FIS into the VISSIM. The FIS was coded by using VisVap, a component of VISSIM, originally designed for creating

signal control algorithms. VisVap is a highly specialized language with limited capabilities for other applications. However, it was possible to write an algorithm in VisVap which describes the operation of a FIS. As discussed earlier, however, the FIS had been already restricted to specific membership functions (trapezoid shape only) because of the technical limitations of VisVap.

- The third part includes data collection (Step 4) and the assessment (Step 5). In Step 4 the simulation is carried out for a four hour period by using a volume distribution which is different from the distribution used during the FIS development stage (Step 1 and Step 2). The collected data (testing data) includes detector occupancies at all three detectors, the measured NVDZ, and the estimated value of NVDZ determined by the FIS defined in Step 2. Step 5 has two general objectives, to assess accuracy and sensitivity, but not all two characteristics are assessed in each test.

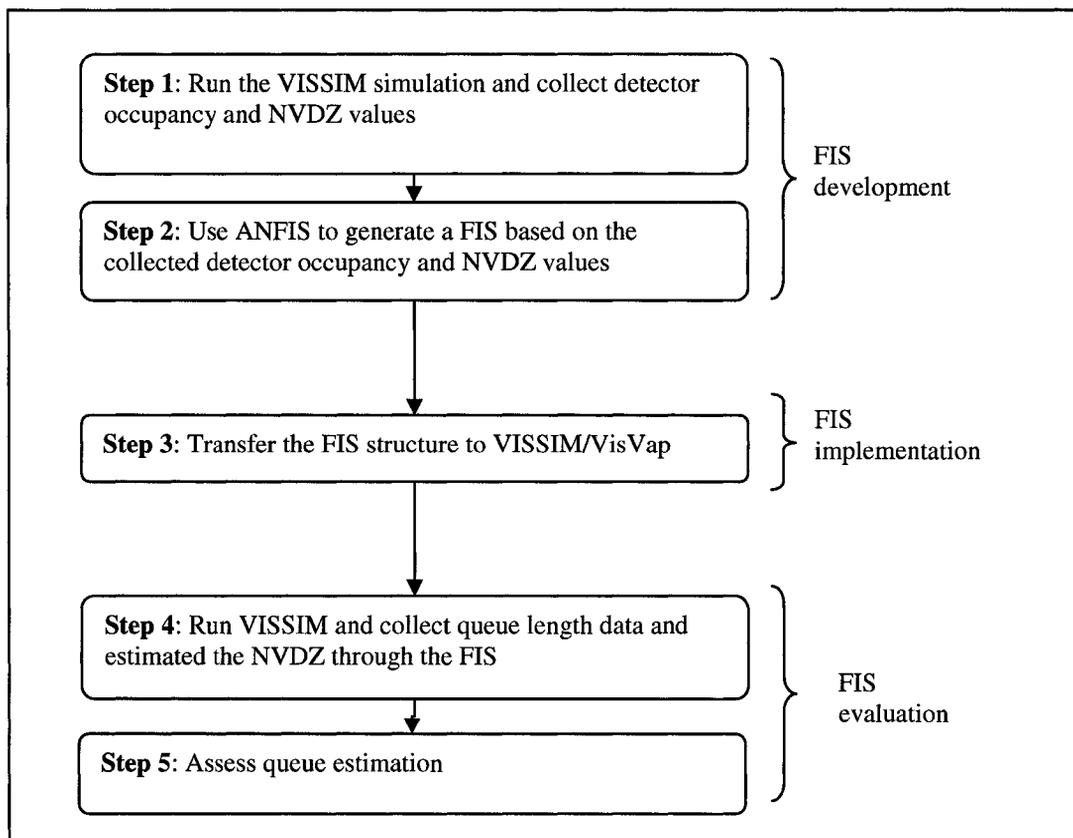


Figure 10.15: Steps for Testing Fuzzy Logic NVDZ Estimation

10.5.1 Evaluation Results

The results of seven tests are reported in this section. The general geometric set-up is depicted in Figure 10.16. Each test is based on a large number of input variables including detection point locations, signal timing, and input volumes. Different tests use different input values therefore the inputs were defined by creating different scenarios. For example, the detection location in Scenario A has 60 metres between occupancy detectors while Scenario D has 40 metres between the detectors.

The input variable scenario used in each test is only noted in this section and a detailed description of the scenario is provided in Appendix T. Exceptions to this approach are made when showing the scenarios as part of the main text helps the reader to follow the discussion.

This section contains only the results of the most important tests. Appendix T includes the MoP for some other, less relevant data sets. The description of all data sets and the corresponding MoP are also provided in the same Appendix. For referencing purposes all data sets are identified by a Data Set #.

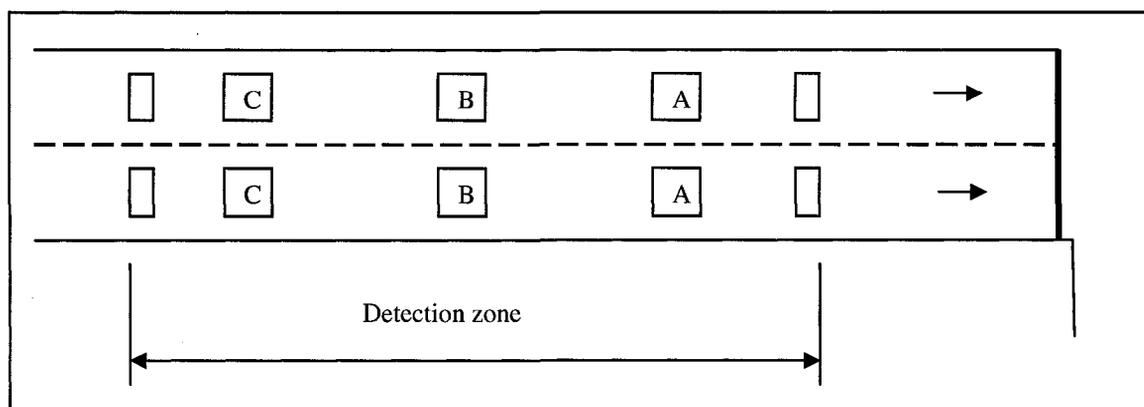


Figure 10.16: General Geometric Set up

The results of seven tests are presented below. The objectives of individual tests are the following:

Test 1: Establish a Base Case Scenario (BCS)

Test 2: Investigate the impact of detection coverage

Test 3: Investigate the impact of the number of lanes

Test 4: Investigate the impact of reduced data points

Test 5: Investigate the location of detection zone relative to the downstream signal

Test 6: Investigate the impact of detection zone length

Test 7: Investigate the impact of different signal timing

Test 1: Base Case Scenario

The geometric set-up and operational characteristics described in Table 10.5 represent the Base Case Scenario (BCS). This scenario is used as the basis for comparison in most of the following tests.

Table 10.5: Test 1 Scenarios

Data Set	Signal Timing Scenario	Detection Location Scenario	Volume Scenario
13	A	A	A

Note: Scenario descriptions are provided in Appendix T.

Based on the input data the measured and estimated NVDZ were obtained from VISSIM and are presented in Figure 10.17. The MoP, describing the accuracy of the estimate, are presented in Table 10.6.

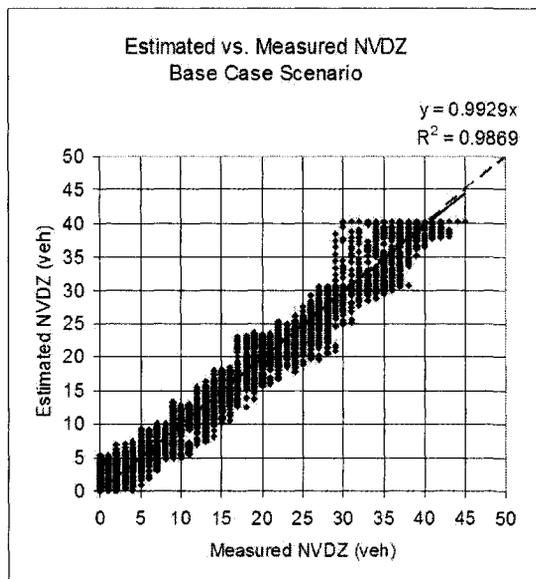


Figure 10.17: Estimated vs. Measured NVDZ for the BCS

Table 10.6: Selected MOP for the Base Case Scenario

Data Set	R	m	RMSE	e	Max E
13	0.993	0.993	1.83	1.28	10.3

The correlation between the estimated and observed NVDZ is obvious and the coefficient of correlation is relatively high; 0.993. The tangent (m) of the least square trend-line (forced to go through the origin) is 0.993 which is very close to 45 degrees and is indicating an excellent overall behaviour for the entire range of observation.

The average absolute error |e| of 1.28 vehicles indicates the overall accuracy of the estimation and the maximum absolute error Max|E| of 10.3 is an indication of the maximum variability of the estimation. Since the NVDZ is measured and estimated for two lanes of traffic, the average and maximum errors apply to two lanes. For one lane, the average error would be less than one vehicle and the maximum error would be

approximately 5 vehicles. The root mean square of the error (RMSE) is also presented in the table, but this will be more useful in the sensitivity analysis when this set up is compared to others.

The absolute error is relatively consistent throughout the range of observation. However, in terms of relative accuracy, the estimation improves as the value of the NVDZ increases. At measured NVDZ of 10 vehicles, the maximum error ($\text{Max}|E|$) is 70% of the measured NVDZ length while at 40 vehicle NVDZ the maximum error is only 10%.

Table 10.7: The Change of Estimation Error as NVDZ Increases

Measured Value (vehicles)	Max $ E $ (vehicles)	Max $ E $ as % of Measured Value
10	7	70
20	7	35
30	11	37
40	4	10

Test 2: Sensitivity to detection coverage

The objective of this test was to investigate the possibility of collecting occupancy data in only one lane of a multi lane roadway with equal lane utilization. Installing detection in only one lane has the advantage to lower the costs of implementation and maintenance.

Table 10.8: Test 2 Scenarios

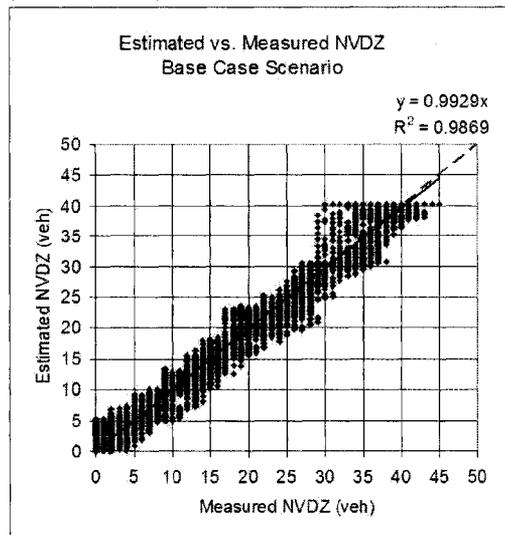
Data Set	Signal Timing Scenario	Detection Location Scenario	Volume Scenario
10	A	A(in one lane)	A

Detection coverage in one or two lanes is independent of the number of detection points used to collect occupancy data. The number of detection points for data collection and

FIS generation is still three. In case of detection coverage in two lanes adjacent detectors are connected and the occupancy values are combined.

The estimated and measured NVDZ for the single lane detection coverage are compared in the right hand graph in Figure 10.18. The left hand graph depicts the data points for the BCS for reference purposes.

Detection in two lanes
(Data Set 13)



Detection in one lane
(Data Set 10)

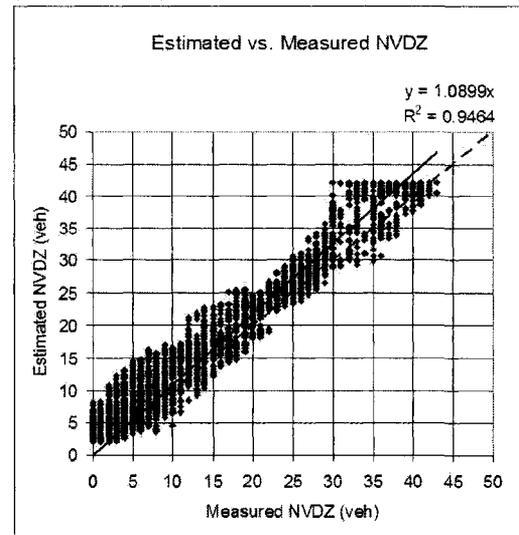


Figure 10.18: Comparison of the NVDZ Estimates for Different Detection Coverage

Table 10.9: MOP for Different Detection Coverage

Scenario	R	m	RMSE	e	Max E
Detection in two lanes (BCS - Data Set 13)	.993	0.993	1.83	1.28	10.3
Detection in one lane (Data Set 10)	0.973	1.089	3.08	2.72	12.3

The comparison of the two plots (Figure 10.18) indicates that the estimates based on detection in both lanes are better when the NVDZ are smaller. For small NVDZ values,

detection in only one lane tends to overestimate the NVDZ. The overall accuracy of the estimation, expressed by the RMSE is approximately 70 % worse (3.08 compared to 1.83), the average absolute error is over 100% higher (2.72 compared to 1.28), and the maximum absolute error is approximately 20 % higher (12.3 compared to 10.3). The difference between the 45 degree line and the tangent of the trend line (expressed by the slope 'm') also indicates that there is an estimation bias within the range of observation. The above analysis indicates that the NVDZ estimation accuracy is rather sensitive to the detection coverage. However, this does not necessarily mean that a set up with detection in both lanes is always preferred. The selection of detection coverage would normally be based on a decision that results in the lowest cost while having adequate accuracy, and the overall accuracy of the single lane detection coverage is still relatively high.

Test 3: Sensitivity to the number of lanes of traffic

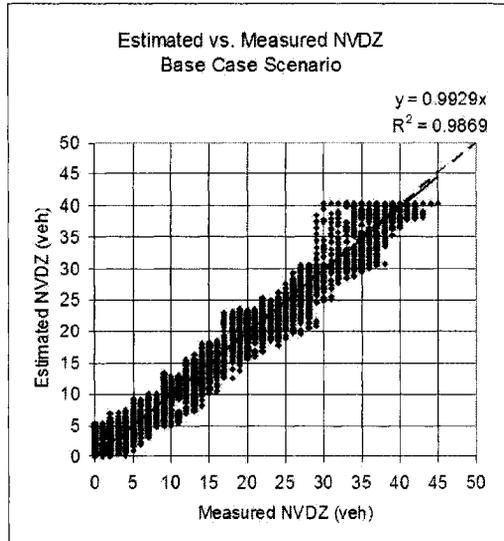
The objective of this test is to investigate whether the NVDZ can be estimated more precisely for roadways with only one lane of traffic. The set up of the detection zone and signal timing remained unchanged therefore the volumes used for testing were only half of the volumes used for the two lane operation scenario (Scenario B in Table 10.10).

Table 10.10: Test 3 Scenarios

Data Set	Signal Timing Scenario	Detection Location Scenario	Volume Scenario
12	A	A	B

Obviously, the maximum value for the NVDZ within the same detection zone lengths is only half compared to two lane traffic which is reflected on the plots shown in Figure 10.19.

Traffic Operation in two Lanes
(Data Set 13)



Traffic Operation in one Lane
(Data Set 12)

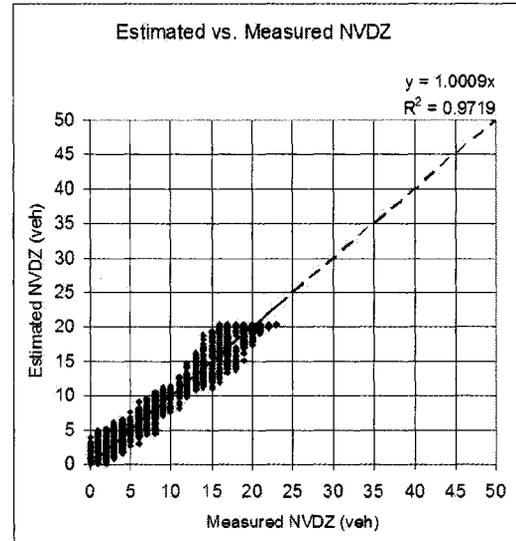


Figure 10.19: Comparison of NVDZ Estimates for Operation on Roadways with Different Number of Lanes

Although the plots indicate a tighter spread of points, the MOP are indicating that the accuracy of the NVDZ estimation for one lane of operation is very similar to the estimates for two lane operation. The coefficients of correlation and the tangents of the trend lines are very similar. The other three MOP are approximately 50% of the corresponding MOP for two lane traffic operation. This is to be expected since the NVDZ for one lane is only half of the NVDZ in the two lanes.

Table 10.11: MoP for Roadways with Different Number of Lanes

Scenario	R	m	RMSE	e	Max E
Traffic operation in two lanes (BCS - Data Set 13)	.993	0.993	1.83	1.28	10.3
Traffic operation in one lane (Data Set 12)	.986	1.001	0.94	0.67	4.8

It can be concluded that the NVDZ estimation is not sensitive to the number of traffic lanes as long as there are detectors in all lanes.

Test 4: Sensitivity to the number of records

This test is somewhat different from the others. Instead of investigating the impacts of geometric set up and/or operations characteristics, it focuses on the effectiveness of the data collection process for FIS development (Table 10.12).

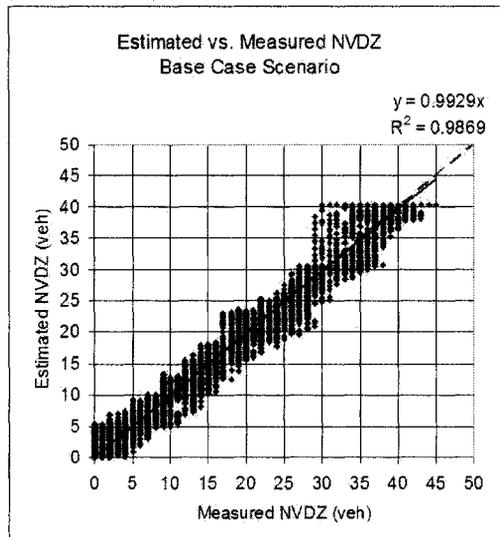
Table 10.12: Test 4 Scenarios

Data Set	Signal Timing Scenario	Detection Location Scenario	Volume Scenario
14	A	A	B (first 7200 records)

The development of the FIS, which is the foundation of the NVDZ estimation method, is based on data collected by simulation. Therefore, collecting sufficient data points is not a major issue. Nevertheless, it is important to know the benefits of having larger data sets.

The FIS developed for the BCS used 14400 data points which is equivalent to 4 hours of simulation. This was compared to a FIS based on 7200 records the equivalent of 2 hours of simulation. The results are presented in Figure 10.20 and in Table 10.13.

FIS based in 14400 records
(Data Set 13)



FIS based in 7200 records
(Data Set 14)

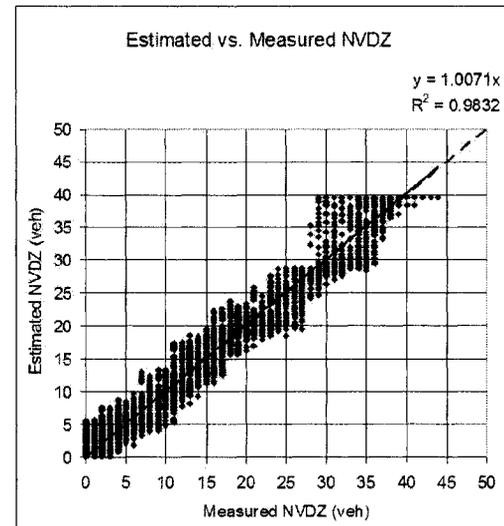


Figure 10.20: Comparison of NVDZ Estimates based on Different Number of Records

The comparison of the NVDZ estimates vs. measurement plots does not suggest that the FIS based on a smaller number of data points had a large negative impact on the NVDZ estimation capability. All MoP, except the $\text{Max}|E|$ were slightly worse for the FIS generated by using less records (Table 10.13). This does certainly not indicate that generating the FIS by using smaller number of points is always better. It is more likely that the first 7200 records included most of the relevant data combinations and the second part of the data set did not contribute much to the development of the FIS.

Table 10.13: MoP for Estimates based on Different Number of Records

Scenario	R	m	RMSE	e	Max E
FIS based on 14400 records (BCS - Data Set 13)	.993	0.993	1.83	1.28	10.3
FIS based on 7200 records (Data Set 14)	.992	1.007	1.86	1.30	10.7

The general conclusion is that the quality of data is more important than quantity. The term quality can be defined as representative. Larger data sets are expected to be better because they increase the chances of containing more relevant records.

Test 5: Sensitivity to the location of detection zone relative to the downstream signal

Queue formation and dissipation within the detection zone is expected to be different as the detection zone is moved closer or further from the signalized intersection. To investigate the general impact of the relative location of the detection zone all detectors were moved closer to the intersection (Scenario B, Table 10.14 and Figure 10.21).

Table 10.14: Test 5 Scenarios

Data Set	Signal Timing Scenario	Detection Location Scenario	Volume Scenario
17	A	B	C

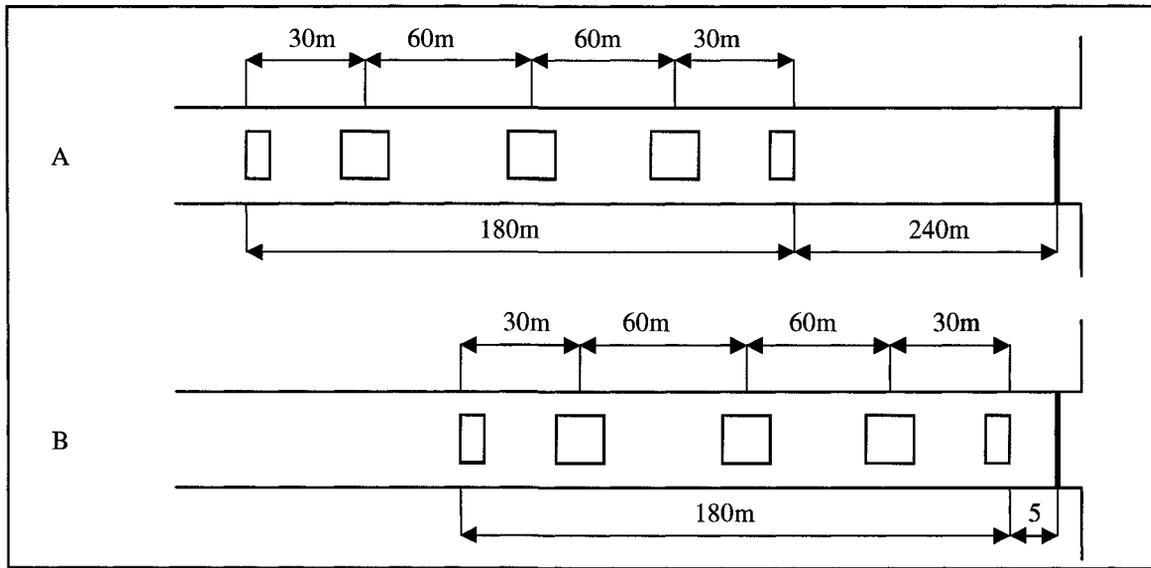
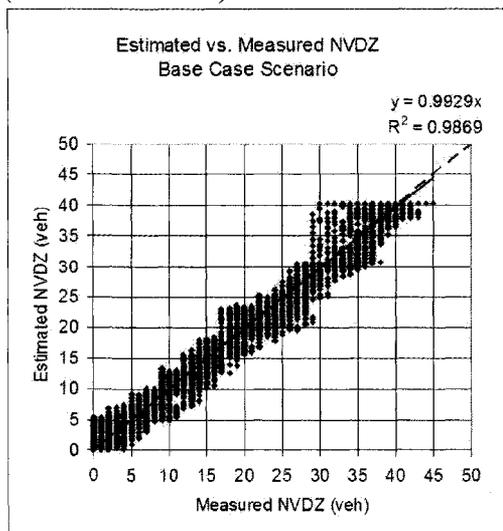


Figure 10.21: Scenarios for Detector Locations

The plot of the estimated vs. measured NVDZ values (Figure 10.22) indicates that as the NVDZ increases there is a gradual underestimation of the NVDZ. The underestimation is about 10 %, thus at a NVDZ of 40 vehicles, the magnitude of the error is approximately 4 vehicles.

Original detector location
(BCS Data Set 13)



Detectors moved towards the signal
(Data Set 17)

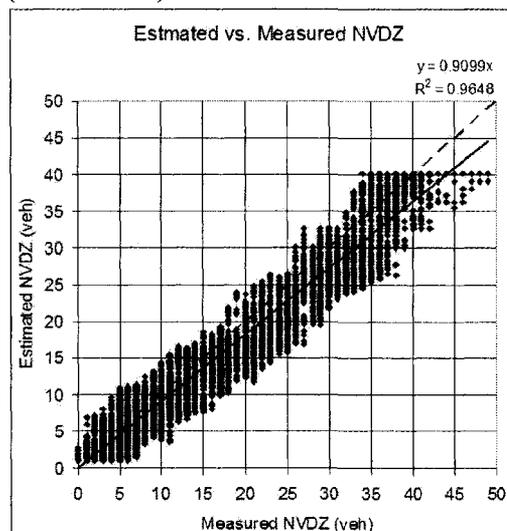


Figure 10.22: Comparison of NVDZ Estimates for Different Data Collection Point Locations

The MoP indicate that the variability of the estimates is larger when the detection zone is closer to the intersection (Table 10.15). The RMSE and the average absolute error $|e|$ are significantly higher by approximately 63 % and 80 % respectively. Finally the maximum absolute error $\text{Max}|E|$ is higher as well, but only by approximately 13% (Table 10.15).

Table 10.15: MoP for Estimates Based on Detector Locations Relative to the Intersection

Scenario	R	m	RMSE	$ e $	$\text{Max} E $
Detection location : Scenario A (BCS - Data Set 13)	.993	0.993	1.83	1.28	10.3
Detection location: Scenario B (Data Set 17)	.982	0.91	2.98	2.31	11.7

The poorer performance of the NVDZ estimation closer to the intersection is to be expected. The impacts of queue formation and dissipation in terms of detector occupancies are much greater at locations closer to the intersection. This conclusion may have some important repercussions on the design of the detection zone.

The tangent of the trend line constructed by using estimated and measured NVDZ is different than the expected 45 degree line. This results in a consistent underestimation of the NVDZ. Further research is required to find the reasons and remove this bias (Section 11.2).

Test 6: Sensitivity to detection zone length

It seems to be logical that as the length of the detection zone increases the accuracy of the NVDZ, by using the same number of detectors, decreases. The question that needs to be answered is to what extent accuracy drops when detection zone length increases?

Two detection zone length scenarios were compared to the BCS. In Scenario D the detection zone length was decreased from 180 to 120 meters. In Scenario E, the length of the detection zone was increased from 180 to 240 meters (Figure 10.23).

Table 10.16: Test 6 Scenarios

Data Set	Signal Timing Scenario	Detection Location Scenario	Volume Scenario
20	A	D	A
21	A	E	A

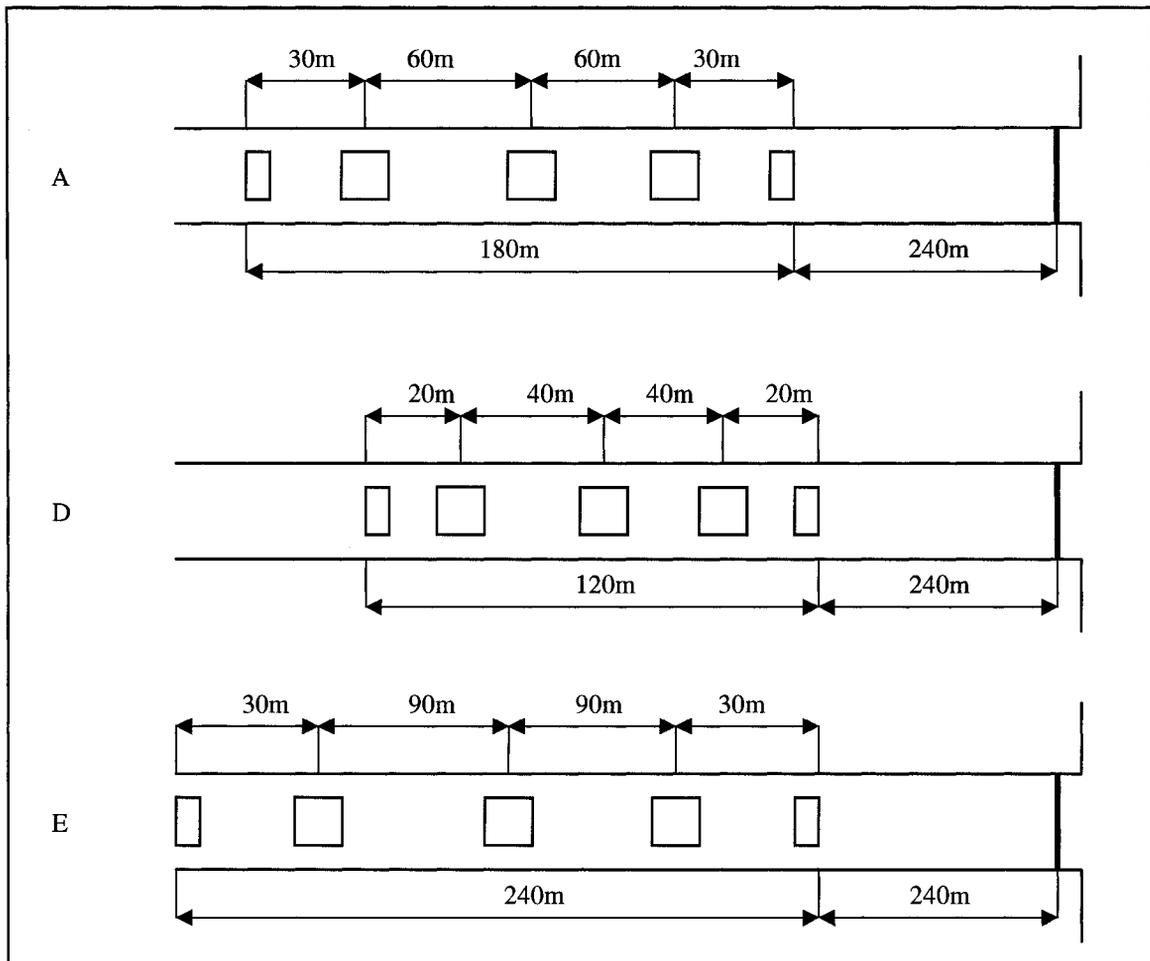
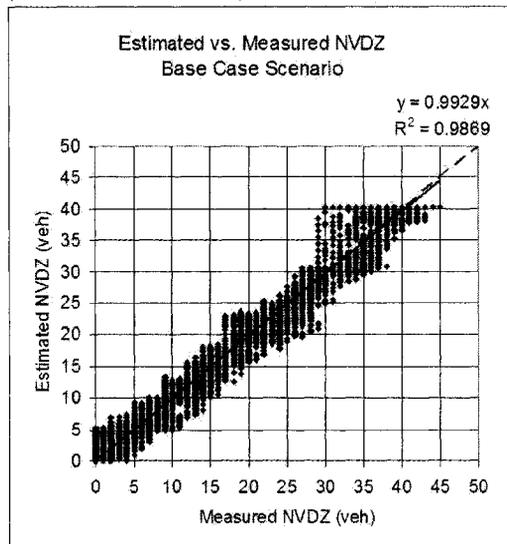


Figure 10.23: Detection Zone Length Scenarios

The correlation between the estimated and measured queue lengths is quite obvious (Figure 10.24) and high for both the decreased and increased detection zone scenarios.

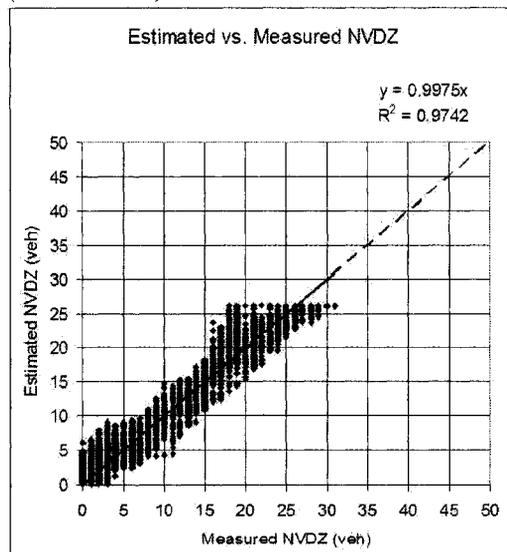
The figure also shows that the maximum number of vehicles within the detection zone is different for the three scenarios.

180 m detection zone, Scenario A
(BCS – Data Set 13)



This space is intentionally left empty

120 m detection zone, Scenario D
(Data Set 20)



240 m detection zone, Scenario E
(Data Set 21)

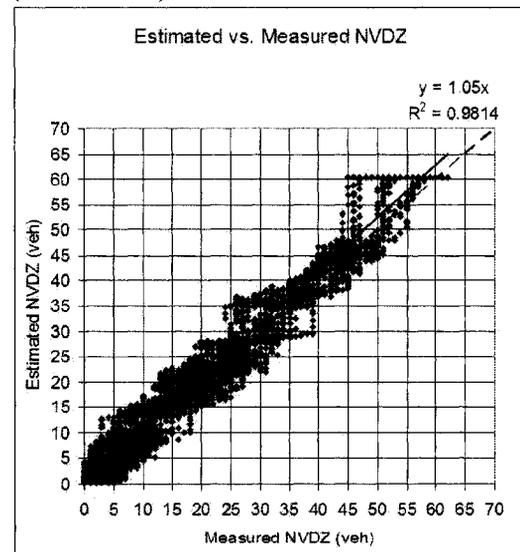


Figure 10.24: Comparison of Queue Length Estimates for Different Detection Zone Lengths

By comparing the results it can be confirmed that estimation accuracy increases with the reduced detection zone length (Table 10.17). All MoP are slightly better when the detection zone is 120 m compared to the BCS with 180 m.

The MoP for the longer detection zone show that the estimation accuracy drops quite significantly as the detection zone length increases. The average absolute error $|e|$ and the maximum error $\text{Max}|E|$ increase by approximately 86% and 50% respectively compared to the BCS.

Table 10.17: MoP for Different Detection Zone Lengths

Scenario	R	m	RMSE	$ e $	$\text{Max} E $
180 m detection zone length, Scenario A (BCS - Data Set 13)	.993	0.993	1.83	1.28	10.3
120 m detection zone length, Scenario D (Data Set 20)	.987	0.998	1.77	1.3	8.2
240 m detection zone length: Scenario E (Data Set 21)	.992	1.05	3.09	2.39	15.5

It is likely that increasing NVDZ estimation accuracy is possible by increasing the number of occupancy detection points from the currently used three per lane to four or more. This would increase the size of the rule base. Four detectors would require the definition of 81 rules (3^4) but the rule generation could be done by using ANFIS without technical difficulties.

Test 7a: Sensitivity to different signal timing (different stage length)

A FIS developed for a specific location is not transferable to a different location since its membership function parameters were tuned by using queue formation and dissipation characteristics for that specific location. However, signal timing has also an impact on the

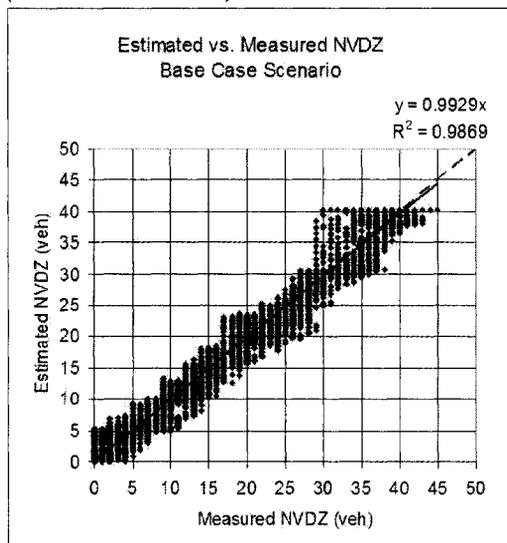
queue formation and dissipation dynamics. Therefore, the performance of the FIS developed for specific signal timing may not be very good when signal timing changes. The objective of this test is to investigate the transferability of a FIS developed for one signal timing to another signal timing at the same location and identical geometric set up.

In the first part of the test, the cycle length was kept constant. The FIS developed for the BCS was implemented for two different signal timing plans. The BCS had a 50 second cycle length and a 15 second green stage for the observed approach (50/15). This was compared to two different timing plans, one with a 20 second green stage (50/20), and the other with a 25 second green stage (50/25)

Table 10.18: Test 7a Scenarios

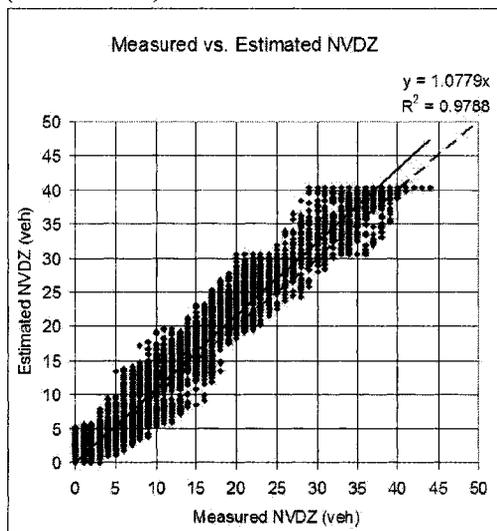
Data Set	Signal Timing Scenario	Detection Location Scenario	Volume Scenario
35	A(50/15)→B(50/20)	A	E
34	A(50/15)→C50/25)	A	D

50/15 timing, Scenario A
(BCS- Data Set 13)



This space is intentionally left empty

50/20 timing, Scenario B
(Data Set 35)



50/25 timing Scenario C
(Data Set 34)

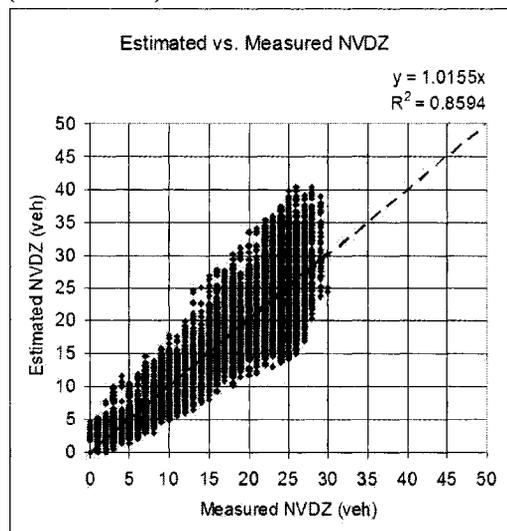


Figure 10.25: Comparison of NVDZ Estimates for Different Stage Lengths

The plots of the estimated vs. the measured NVDZ reveal that an increase of the stage length has a negative impact on the NVDZ estimation. This is particularly evident when

the stage length was increased to 25 seconds (bottom right plot on Figure 10.25). The figure also reveals that the maximum NVDZ is by approximately 10 vehicles smaller than the maximum NVDZ on the other two graphs. The reason for this is that with a 25 second stage length, vehicles are always moving through the entire part of the detection zone, even when the downstream signal is red. Thus there is always more space between vehicles than when the queue totally stops for certain time periods.

The MoP (Table 10.19) also reflect the poorer performance of the queue estimation with the longer stage lengths.

Table 10.19: MoP for Different Stage Lengths

Scenario	R	m	RMSE	e	Max E
50/15 timing, Scenario A (BCS – Data Set 13)	.993	0.993	1.83	1.28	10.3
50/20 timing, Scenario B (Data Set 35)	.989	1.078	3.05	2.31	10.9
50/25 timing, Scenario C (Data Set 34)	.927	1.016	3.69	2.74	14.9

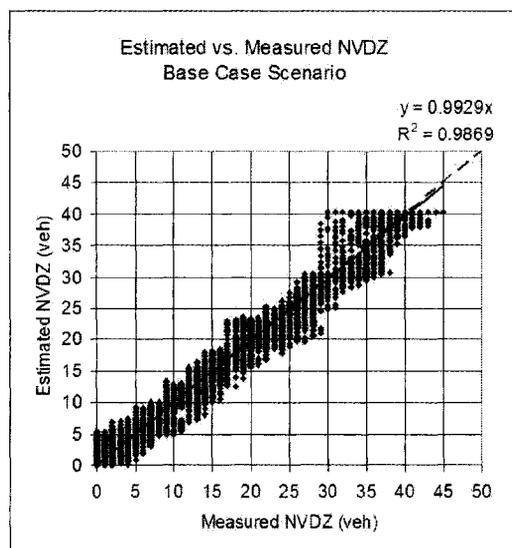
Test 7b: Sensitivity/transferability to different signal timing (different cycle and stage time)

The second part of the test investigates the transferability of a FIS developed for an intersection with one cycle length to signal operation at the same intersection with a different cycle length. The FIS developed for the BCS with a 50 second cycle length and 15 second stage duration (50/15) was applied to operation with signal timing with a 130 second cycle length and 60 seconds stage duration (130/60).

Table 10.20: Test 7b Scenario, Part 1

Data Set	Signal Timing Scenario	Detection Location Scenario	Volume Scenario
33	A(50/15)→D(130/60)	A	D

FIS developed for 50/15 signal timing
(BCS - Data Set 13)



FIS applied to 130/60 signal timing
(Data Set 33)

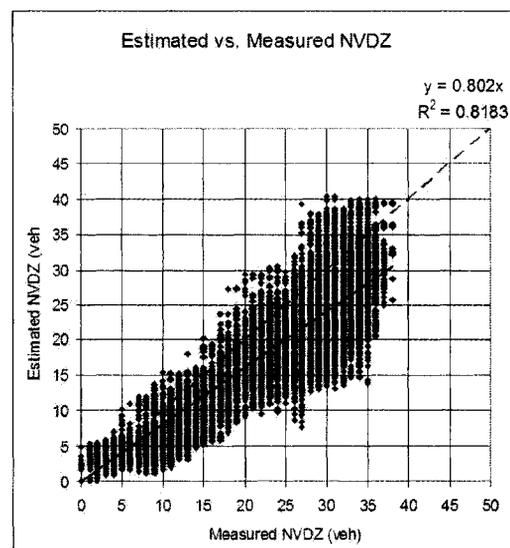


Figure 10.26: The Application of FIS Developed for Signal Timing 50/15 to Signal Timing 130/60

The comparison of the estimated vs. measured NVDZ plots (Figure 10.26) shows that by transferring a FIS developed for one signal timing plan to an intersection with a different signal timing plan, the accuracy of the NVDZ estimation drops substantially. The analysis of MoP (Table 10.21) also shows the dramatic loss of the FIS's ability to provide a good NVDZ estimate for different signal timing.

Table 10.21: MoP for Different Cycle Length, Part 1

Scenario	R	m	RMSE	e	Max E
FIS developed for 50/15, Scenario A (BCS - Data Set 13)	.993	0.993	1.83	1.28	10.3
FIS applied to 130/60, Scenario D (Data Set 33)	.904	0.802	6.3	5.05	21.1

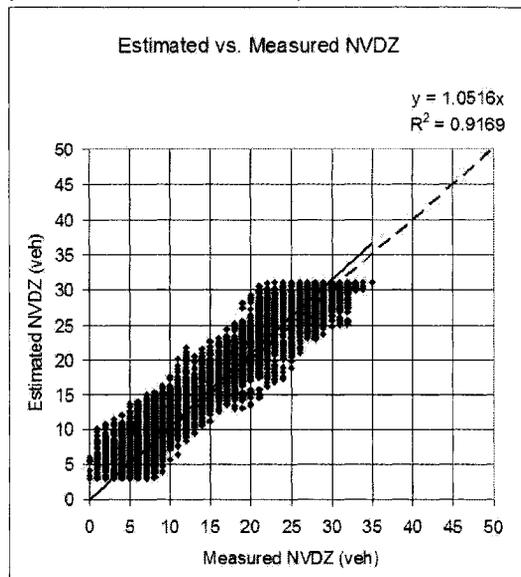
To examine the performance of a FIS developed for longer cycle lengths in an application with shorter cycle length, the previous test was repeated but now the FIS was developed for signal operation with 130/60 timing and applied to signal operation with 50/15 timing.

Table 10.22: Test 7b Scenario, Part 2

Data Set	Signal Timing Scenario	Detection Location Scenario	Volume Scenario
30 NEW BCS	D(130/60)	A	D
32	D(130/60)→A(50/15)	A	A

The first plot (Figure 10.27) shows the new BCS. The FIS for the NVDZ estimation was developed by using 130/60 signal timing. The correlation of the estimated and measured NVDZ is high and the variability is very consistent. If the same FIS is applied to signal timing 50/15, the correlation between the estimated and measured NVDZ becomes markedly worse (second plot in Figure 10.27).

FIS developed for 130/60 Signal timing
(New BCS - Data Set 30)



FIS applied to 50/15 Signal timing
(Data Set 32)

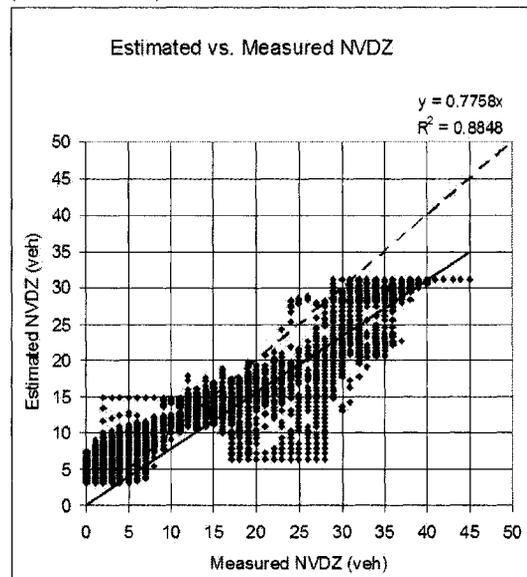


Figure 10.27: The Application of FIS Developed for Signal Timing 130/60 to Signal Timing 50/15

The MoP also indicate a significantly poorer performance for the example with modified signal timing (Table 10.23). All measures are substantially worse.

Table 10.23: MoP for Different Cycle Length, Part 2

Scenario	R	m	RMSE	e	Max E
FIS developed for 130/60, Scenario D (New BCS - Data Set 30)	.958	1.052	2.86	2.31	9.66
FIS applied for 50/15, Scenario A (Data Set 32)	.94	.776	6.16	4.94	21.7

The test demonstrated that the NVDZ estimates were quite reasonable when the FIS was used for same signal timing they were developed for. It is also evident that appropriate FIS could be developed for both a 130 and a 50 second cycle.

To identify the differences between the two FIS, membership function parameters were compared for the two models. The comparison of the input membership function parameters revealed that they are reasonably similar (Table 10.24).

Table 10.24: Membership Function Parameters for Data Sets 13 (Developed for 50/15 Signal Timing) and Data Sets 30 (Developed for 130/60 Signal Timing)

Linguistic variable	Parameter of trapezoid MF	Data Set	
		#13	#30
OccA_S	a	-99	-99
	b	-95	-95
	c	.177	.17
	d	.404	.460
OccA_M	a	.172	.13
	b	.377	.36
	c	.749	.83
	d	.991	.94
OccA_L	a	.569	.50
	b	.891	.85
	c	95	95
	d	99	99
OccB_S	a	-99	-99
	b	-95	-95
	c	.115	.15
	d	.442	.41
OccB_M	a	.03	.11
	b	.278	.35
	c	.656	.65
	d	.978	.81
OccB_L	a	.573	.66
	b	.84	.84
	c	95	95
	d	99	99
OccC_S	a	-99	-99
	b	-95	-95
	c	.119	.18
	d	.409	.56
OccC_M	a	.024	.11
	b	.328	.37
	c	.687	.63
	d	.990	.87
OccC_L	a	.605	.51
	b	.872	.82
	c	95	95
	d	99	99

However, the comparison of the output membership function parameters (which are actually singleton outputs) reveals the differences (Table 10.25). The FIS developed for 130/60 signal timing has lower maximum values for Rules 25, 26, and 27. These are the rules which describe the highest number of vehicles within the detection zone. By observing the simulation it has been noticed that vehicles may move slowly within the detection zone during the red signal but they stop rarely even when there are left over queues. As a result the maximum number of vehicles within the detection zone is lower compared to the situation when vehicles have to stop (Data Set 13).

The second major difference is that Rules 15 to 18 do not have any influence in case of Data Set 13 (outputs are zero), but they play a role in the estimation for Data Set 30.

Based on the above it can be concluded that the data points collected for the development of the FIS for one signal timing may not contain sufficient representative data points for a different signal timing. The FIS can perform well only in a queuing regime for which it has been trained.

Training FIS for different signal timings is a very interesting area to be explored, but outside of the scope of this thesis. For the purpose of this research, NVDZ estimation by assuming relatively stable signal timing is adequate since in oversaturated conditions the signal timing of bottleneck intersections is not highly variable.

Table 10.25: Output Membership Function Values for Data Sets 13 (Developed for 50/15 Signal Timing) and Data Sets 30 (Developed for 130/60 Signal Timing)

Rule	Rule	Data Sets	
		#13	#30
1	SSS	.06	5.9
2	SSM	12.5	17.9
3	SSL	0	0
4	SMS	0	0
5	SMM	10.1	7.2
6	SML	0	0
7	SLS	0	0
8	SLM	0	0
9	SLL	0	0
10	MSS	2.24	5.4
11	MSM	5.26	13.8
12	MSL	0	0
13	MMS	2.96	6.5
14	MMM	10.5	17.7
15	MML	0	18.9
16	MLS	0	36.9
17	MLM	0	10.2
18	MLL	0	24.3
19	LSS	7.3	14.8
20	LSM	11.3	3
21	LSL	0	0
22	LMS	16.9	11.4
23	LMM	21.6	23.7
24	LML	30	31
25	LLS	19.9	6.3
26	LLM	30.7	27.4
27	LLL	40.5	31.1

10.6 Discussion on Using ANFIS for Determining the NVDZ

The relationship between detector occupancies and the corresponding number of vehicles in the detection zone (NVDZ) can be described relatively precisely by a Fuzzy Inference

System on an artificial example. The average accuracy of the estimate is within a few vehicles (<2) per lane with maximum errors of up to 5 vehicles per lane. The NVDZ estimation is sensitive to detection coverage and to the length of the detection zone.

In terms of transferability of a FIS developed at an intersection approach with one timing plan to the same intersection approach with a different timing plan, more research is needed. Based on the evaluation in this research, NVDZ estimation errors increase dramatically when signal timing changes. This is however, not an issue for the intended purpose of the NVDZ estimation technique developed in this section. In oversaturated conditions, signal timing changes are normally relatively small.

10.7 Simulation of Real-World Example: Queue Management with Fuzzy Logic Based Estimation of NVDZ

10.7.1 Geometric Set-Up and Operations Description

Roadway geometry, signal operation, and input volumes are identical to the example described in Section 9.4. For simulation purposes, the detector locations are identical. However, the detection requirements in real-world implementation are different. The check-in/check-out detectors are needed only for tuning purposes in a simulation environment and are not required for a real-world application.

10.7.2 Simulation Process

The simulation process for this example differs from the process described in Section 9.4 only in terms of the method used to determine the NVDZ. Previously, the NVDZ were determined by counting vehicles as they enter and exit the detection zone. In this case, however, the number of vehicles is estimated by using vehicle occupancy data at three data collection points as described in Section 10.4.

The simulation process has two distinct parts. In the first part, the parameters of the FIS for estimating the Number of Vehicles in the Detection Zone (NVDZ) are determined. This process has three steps as described in Figure 10.28. These steps were presented in Section 10.5 and they will be discussed here only briefly.

The data collection through VISSIM (Step 1) is carried out by running the simulation for the specific physical set-up and input data. During this step, the signal at the metering intersection is operating without the queue management algorithm. The objective is to collect a set of data that describes the relationship between detector occupancies and the corresponding NVDZ in this specific environment. In Step 2 FIS, parameters are generated by using MATLAB's ANFIS tool, and in Step 3, the FIS structure is coded into VISSIM/VisVap.

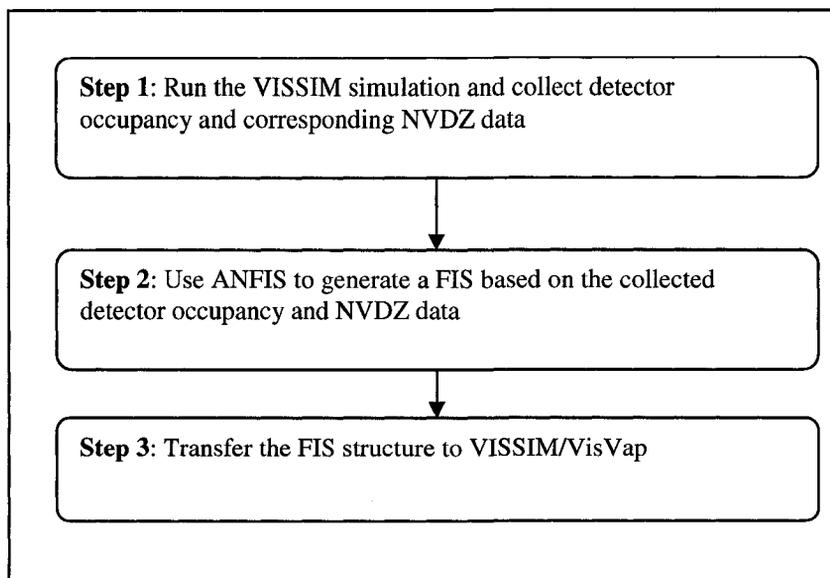


Figure 10.28: FIS Parameter Tuning for Estimating the NVDZ

The second part of the simulation process includes a VISSIM run by applying the queue management algorithm described in Figure 10.29. This algorithm is very similar to the one described in Figure 9.5. The only difference is in the method of determining the number of vehicles in the detection zone.

The simulation was carried 20 times with different random number generating seeds and each time the simulation lasted for a 2 hour period. The presented MOP are based on the average values of the 20 samples.

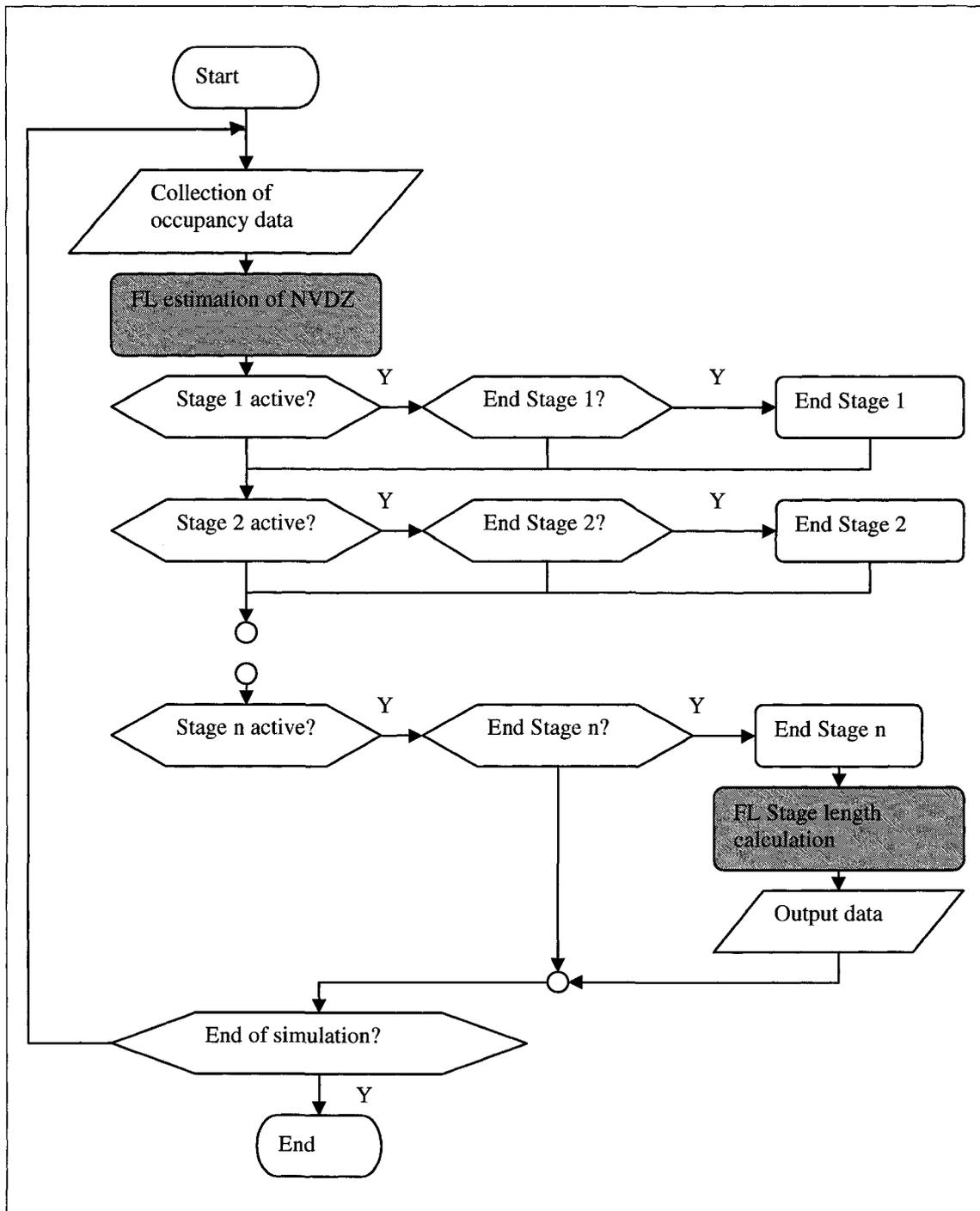


Figure 10.29: Queue Management Controller Algorithm Based on FL NVDZ Estimation

10.7.3 Evaluation of Simulation Results

The objective of evaluation is to investigate whether the Fuzzy Logic based estimation of the number of vehicles in the detection zone (NVDZ) can be used instead of relying on vehicle counts collected with check-in/check-out detectors.

The comparison of MOP is carried out on the aggregate and disaggregate levels. On the aggregate level, the mean travel times and total person travel times are compared. On the disaggregate level travel times on individual travel sections were examined.

In terms of the mean vehicle delay, the difference between the two methods is very small and statistically insignificant at the 95% confidence level (Table 10.26). Similarly, the difference in total person delay is insignificant as well (Table 10.27).

Table 10.26: Comparison of Mean Travel Times

		Mean Vehicle Travel Times (sec)		
		NVDZ based on counts (sec)	NVDZ based on FL estimate (sec)	Difference (sec)
Car	Mean	280.0	280.8	0.8 (0.3%)
	σ	35.6	35.7	
Bus	Mean	213.2	212.9	-0.3 (-0.1%)
	σ	1.7	1.8	
Car+Bus (Total)	Mean	277.6	278.4	0.7 (0.3%)
	σ	34.3	34.4	

Table 10.27: Comparison of Total Person Travel Times

		Total Person Travel Times (pers-hr)		
		NVDZ based on counts (pers-hr)	NVDZ based on FL estimate (pers-hr)	Difference (sec)
Car+Bus	Mean	828.3	828.5	0.2 (0.02%)

Travel times generated by the two algorithms were also compared on a disaggregate level, for each travel section separately. Results are presented visually on Figure 10.30. None of the travel time differences are statistically significant at the 95% confidence level. (Actual travel time data and the statistical analysis are summarized in Appendix T.)

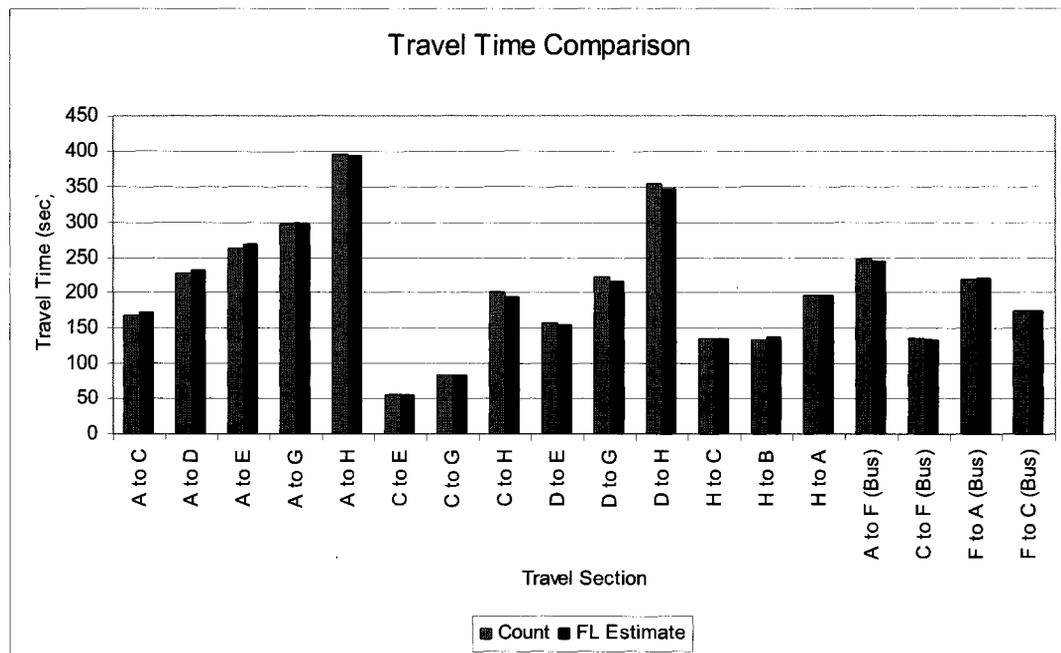


Figure 10.30: Comparison of Travel Times for the two NVDZ Calculation Methods

The analysis at both aggregate and disaggregate levels show that the performance of the two queue management algorithms is very similar. Thus, the Fuzzy Logic based estimation of the number of vehicles in the detection zone based on detector occupancy data is a feasible alternative to detector counts.

10.7.4 Analysis of the Control Algorithm Performance

As the final step in the comparison process the controller outputs and the corresponding NVDZ values are analyzed. The duration of green times and its oscillation is very similar for both algorithms (Figure 10.31, Figure 10.32). The NVDZ values are also very similar except that on the second Figure, the NVDZ values are systematically higher during the simulation time between 0 and 3600 seconds. The reasons for these higher values is that the FL based NVDZ estimation, as tuned for this example, overestimates the number of vehicles when the number of vehicles within the detection zone is small (The occupancy of detector A is small but the occupancy of the other two detectors is close to zero). During the simulation period from 6100 to 7200 seconds the FL based NVDZ estimate goes back to zero because at this time the queue has dissipated and there are very few vehicles within the detection zone (The occupancies of all three detectors are close to zero).

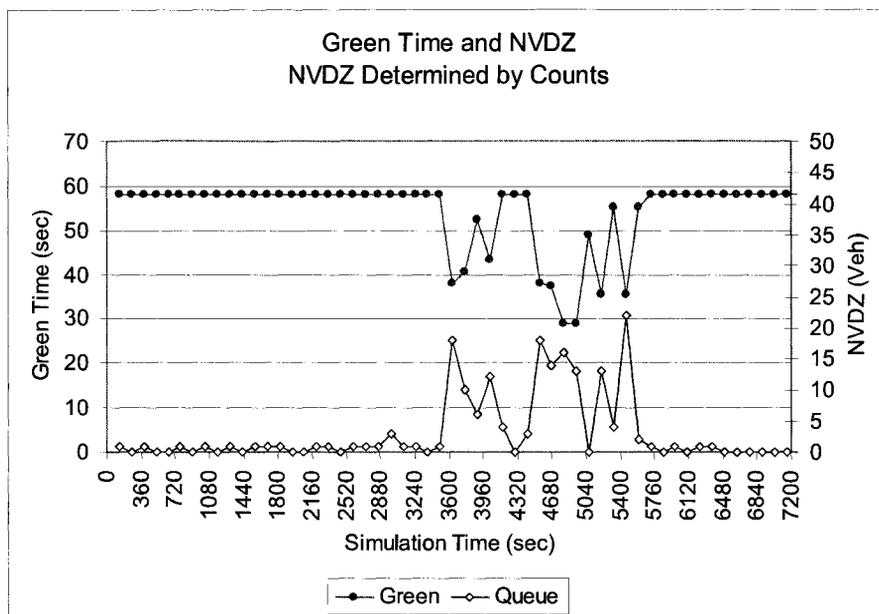


Figure 10.31: Green Time and NVDZ Variation when the NVDZ is Determined by Using Vehicle Counts

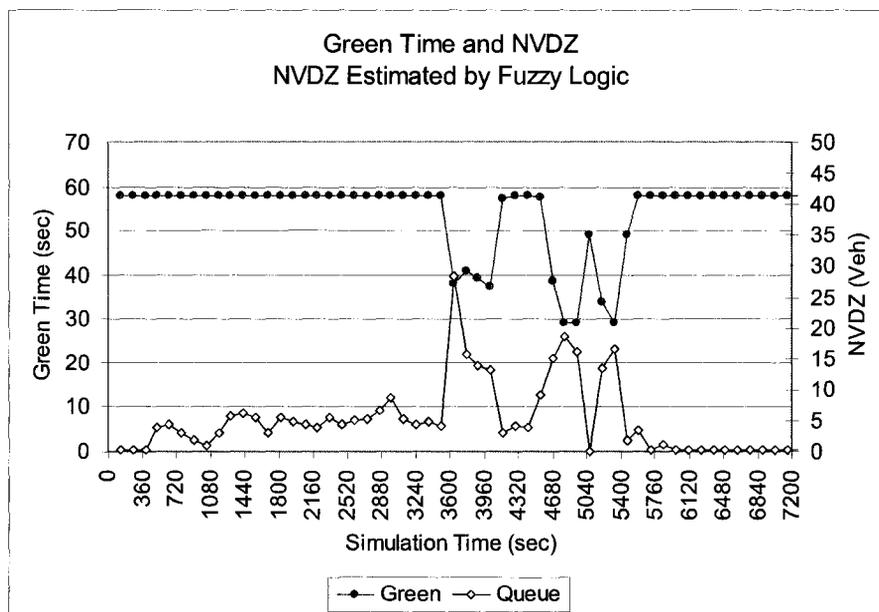


Figure 10.32: Green Time and NVDZ Variation when the NVDZ is Determined by Fuzzy Logic Estimation Based on Occupancy Data

10.8 Summary

The queue management techniques presented in previous chapters require some mechanism to estimate the number of vehicles within a specific section of the roadway. The method of counting vehicles with check-in/check-out detectors has serious practical limitation. Any consistent over or undercounting eventually generates large cumulative errors which decreases and eventually eliminates the effectiveness of any queue management algorithm. To facilitate the estimation of the number of vehicles in a detection zone (NVDZ), a method based on detector occupancies is proposed. The relationship between detector occupancies and the corresponding NVDZ is described by a Fuzzy Inference System, and the membership functions are tuned by using an Adaptive Neuro-Fuzzy Inference System.

The evaluation of proposed NVDZ estimation process shows that using a FIS generated and tuned by ANFIS provides a feasible method to estimate the number of vehicles in the detection zone based on vehicle occupancy data.

The requirement to estimate the number of vehicles in a specific detection zone is not unique to the queue management approach developed in this thesis. A similar measurement is likely needed in any real-time queue management strategy. The Fuzzy Logic based method for estimating the number of vehicles within a detection zone based on detector occupancy data, as proposed in this section, would be useful in other applications as well.

11. FINAL DISCUSSION, CONCLUSION, AND RECOMMENDATION

11.1 Discussion and Conclusions

Since roadway capacity in built up urban areas cannot be increased sufficiently to keep pace with the growth of personal vehicle travel demand, congestion develops and transit vehicles are forced to travel in congested conditions on most of the transportation network. To increase transit modal share in this operational environment bimodal transportation management is required.

Transportation management in oversaturated conditions is obviously not a simple problem, it has both technical and philosophical dimension. The philosophical dimension relates to the general approach to solving the urban transport problem. The technical dimension pertains to the development of queue management techniques coupled with transit priority measures. Furthermore, a new evaluation framework is required which goes beyond the short-term- single mode focus of existing Measures of Performance and is designed to estimate the short and long-term benefits of bimodal (car and transit) traffic management.

11.1.1 Changing the Philosophical Environment of Transportation Planning and Traffic Control

The existing philosophical framework of transportation planning and operation is based on the following three assumptions:

- there is no limitation to vehicle capacity increase;
- the unimportance of treating transit as a system; and
- the independence of long-term planning and short-term operations objectives.

To be able to develop an effective traffic management concept for congested transportation environments these three assumptions and the resulting philosophical frameworks have to be scrutinized and adjusted (Table 11.1).

Table 11.1: Existing and Proposed Philosophical Framework

Existing		Proposed	
Assumption	Philosophical Framework	Assumption	Philosophical Framework
Limitless vehicle capacity	Predict and provide	Limited vehicle capacity	Predict and adapt
	Independence of the elements of the transportation infrastructure		Interdependence of the elements of the transportation infrastructure
	Absence of trade-offs		Analysis of transit system operation
Transit vehicles operate independently (similarly to cars)	Analysis of transit vehicle operation	Transit vehicles are part of the transit system	Analysis of transit system operation
No connection between operations and planning objectives	Independence of operations and planning	Impact of operations on planning objectives	Holistic approach to operations and planning

Limitations to capacity increase

The underlying assumption of most traffic engineering and transportation planning procedures is that there is no long-term limitation to the increase of vehicular capacity. Financial, social, and physical constraints are often believed to be only temporary difficulties which will eventually diminish. The assumption of limitless capacity provided the required background for the development of the “predict and provide” approach for justifying investment in transportation infrastructure. However, in most urban environments it is evident that the limitless capacity increase assumption is not correct any longer and the provision of sufficient vehicular capacity is technically, socially, financially, and environmentally unsustainable. Thus the “predict, determine limitations, and adapt” philosophical approach is proposed (Table 11.1).

Another implication of the limitless capacity increase assumption is the almost exclusive focus on the development of analysis techniques that deal with individual elements of the transportation infrastructure. By assuming that vehicle capacity increase has no firm limits, saturation and oversaturation can be regarded as only temporary problem, thus the need to develop design, evaluation, and operational procedures for saturated and oversaturated conditions is not a priority. Furthermore, under this assumption the traditional traffic engineering philosophy of focusing on individual elements (i.e. intersections) of the transportation infrastructure mostly independently is practical and appropriate. However, if the limitless capacity assumption is not valid and operating under congested conditions becomes a norm rather than an exception, the elements of the

transportation infrastructure cannot be designed, analysed, and operated independently from each other. A new philosophical framework which recognizes the interdependence of the elements of the transportation infrastructure is needed (Table 11.1).

Finally, the limitless capacity increase assumption eliminates the possibility of having contradicting objectives. Theoretically, all objectives in terms of transit, vehicle, pedestrian, and cycling operation could be achieved. Traffic engineering and transportation planning does not have to deal with the difficult concept of trade-offs. However, limitless capacity assumption is false and the unavoidability of compromises is obvious. Nevertheless, the traditional traffic engineering and transportation planning philosophy is ill prepared to deal with trade-offs.

Treating transit as a system

Traditional traffic engineering views transit vehicles similarly to personal vehicles except that they take up more space and green time. Traditional transportation planning looks at transit vehicles from the person carrying capacity perspective. What is common to both traditional approaches is that they treat transit vehicles as independent elements of the transportation system, similarly to personal cars. This is evident by using the mean transit travel time for describing transit operation. However, the operation of individual transit vehicles cannot be analysed in isolation. A philosophical approach which recognizes that transit vehicles are part of the transit system is required. One element of this approach is to use MoP that include the mean transit travel time plus its variability.

Long-term planning and short-term operations objectives

Most long-term transportation plans in Canadian cities place an increased importance on transit. However, to ensure that long-term planning objectives are materialized another shift in the traffic engineering and transportation planning philosophy has to occur. Currently there is a huge disconnect between long-term transportation planning and short-term operations objectives. Long-term plans specify the importance of increasing transit modal share, but the infrastructure design and operation is based on traditional practices which focus on short-term operation efficiency that favour car travel due to the “gaining efficiency through abusing efficiency paradox”. A very significant change has to occur in the philosophy of traffic engineering and transportation planning so that the transportation problem is viewed from a holistic perspective and the links between short-term actions and long term planning objectives are recognized. This link can be established by using identical MoP at the Planning, Design, and Operations levels of decision making. The use of the Transit Car Travel Time ratio (TCTTr) is proposed in this thesis.

11.1.2 Development of Bimodal Traffic Management Techniques

The bimodal traffic control techniques developed in this thesis are founded on the following idea: If intersections represent the bottlenecks on a transportation network then there is some extra capacity on the link connecting intersections. By using this extra capacity of the links to change the sequence of high and low occupancy vehicles, the effectiveness of the system, in terms of total person travel time, can be improved while the vehicular capacity at the bottleneck remains unchanged.

Two distinct queue management techniques were developed and tested, one for operating a partial queue jump, and the other for metering traffic into a bottleneck roadway section. The general signal control algorithms for the two are to a great extent similar.

Four queue management signal control algorithms were developed and evaluated. The simplest algorithm is the “On-Off” control, and it is applicable when queue management is the only function of the controlling signal. The “Park View” type control was developed by modifying a published control algorithm developed in the UK. A model based control algorithm was also developed for testing and comparison purposes. It performed very similarly or even better than the other controls, but it is much more demanding in terms of input data. Finally a Fuzzy Logic based control algorithm was developed. Although in terms of performance measures the Fuzzy Logic control was very similar to the other algorithms, it is a promising approach. Fuzzy Logic control is rule based, which can be relatively easily modified. The general logic is relatively simple, although the tuning of fuzzy logic control systems is still a less formal process.

The four control algorithms were first developed and tested on two models and then implementation was demonstrated on two real-world examples through simulation. One of the real-world examples has been partly implemented in the field but without the use of the control algorithms developed in this thesis.

The analysis and simulation of bimodal oversaturated conditions is significantly different from the traditional single mode uncongested scenario. First, the geographical scope of

the analysis is problematic. In oversaturated conditions queues could easily creep out of any practical scope. The thesis proposes methods to deal with queues creeping both downstream and upstream from the pre-defined scope of analysis. Second, bimodal queue management analysis requires more data than traditional control. Instead of using average hourly traffic volumes, arrival volume distribution data is required.

In a simulation environment data collection is not a major issue. However, in a real-world environment without robust and accurate data collection queue management algorithms cannot operate effectively. Every queue management algorithm discussed in this thesis is based on the number of vehicles approaching or queued up at a particular intersection. This quantity was named NVDZ (number of vehicles in detection zone). Determining the NVDZ can be easily carried out in a simulation environment by counting vehicles with check-in and check/out detectors. However, due to the very likely event of consistent over or undercounting, the cumulative error would render this approach ineffective in real-world applications. Therefore, a method for determining the NVDZ was developed and tested which is based on detector occupancy values instead of detector counts. A fuzzy inference system was developed which maps detector occupancy values into NVDZ value and the Adaptive Neuro Fuzzy Inference System (ANFIS) was used to tune parameters of the fuzzy system.

11.1.3 Bimodal Performance Measurement

The main question is how to measure success and failure in a bimodal environment. If traditional traffic engineering MoP (which have a single mode perspective) are used to assess bimodal transportation management techniques, no bimodal traffic management techniques can be justified.

The MoP have to be selected and the evaluation process must be designed so that the “gaining efficiency by abusing efficiency paradox” and the “propagation of inefficiency trap” are avoided.

The “gaining efficiency by abusing efficiency paradox” describes a situation when a system can be made more efficient by “abusing” its most efficient elements. Translated to the transportation problem this means that the most efficient utilization (i.e. minimum duration of queues and vehicle delay) of the limited roadway space can be achieved by ensuring that low occupancy vehicles which are the least efficient in terms of space utilization, leave the system first and higher occupancy vehicles leave the system last (i.e. “abuse” high occupancy vehicles).

Optimization techniques using performance measures that are based on vehicle delay measurements for existing conditions fall into the “propagation of inefficiency trap”. The trap describes the situation where the inefficient element, since it is preferred by operational measures, inevitably starts to propagate. However, this propagation is not

captured by the optimization or evaluation process and it is not measured by traditional MoP.

To be able to justify bimodal traffic control/management measures, new MoP and evaluation framework are required. The most effective MoP are those that can be related directly to the high level policy/planning and lower level operational objectives of the evaluated transportation system, and which are relatively simple to calculate and/or measure. The selection of MoP depends on the perspective of the evaluation. The following four perspectives were identified:

- Single mode (traditional traffic engineering) perspective
- Bimodal transportation system perspective
- Transit system operations perspective
- Transportation policy/planning perspective

There is no one single MoP that can be used in all above perspectives. Mean vehicle travel time is appropriate for the single mode perspective in oversaturated conditions and it also serves as the basis for person delay calculation which is a more meaningful MoP from the bimodal perspective.

From the transit system operations perspective, however, the mean vehicle travel time is not too helpful; a MoP which includes travel time variability must be used. The use of the 85th percentile travel time is suggested because it better reflects the impact of bimodal traffic management on the possibility of changing transit schedule times.

From the policy/planning perspective the TCTTr which describes the relative advantage of transit travel compared to car travel is proposed. Using TCTTr has a number of advantages. First, it explicitly captures the important factors that determine transit modal share (TMS). Second, measuring the TCTTr is relatively straightforward. It can be carried out for one intersection, a short roadway section or an entire arterial by simulation and/or field measurements. Third, the TCTTr can be directly incorporated into facility planning, design, and operation decisions. For example, if the planning objective is to achieve a 30% TMS in a particular corridor, the required TCTTr can be estimated, and the infrastructure design and operations control/management could be designed accordingly.

11.2 Recommendations for Future Research

This thesis provides sufficient philosophical and technical background to start (continue) with the implementation of the proposed bimodal traffic control techniques. However, there are number of areas that need further research and development.

The controller algorithms, developed in the thesis, need to be implemented in signal controllers. The queue management algorithm requires the introduction of new parameters into control. These new parameters include the selection of metered and compensating stages, and the minimum and maximum times for the metered stages. This work is however, mostly at the implementation and not research level.

In the algorithms developed for this thesis, only one compensating stage was selected. This approach is sufficient for many applications, however at some locations it may be more useful to distribute the extra green time between more competing stages. Further research and development is needed to develop criteria and algorithm for this redistribution. The fuzzy logic approach would seem to be quite effective in this regard.

The thesis focused on the accuracy of control algorithms, but did not explore the need and benefit of responsiveness. Focusing on the accuracy of queue management control is sufficient in many applications because the target queue length (or NVDZ) is often a constant value. However, there are examples where the target queue length (or NVDZ) could change. For example, as a transit vehicle approaches an intersection, the target queue length could be shorter than otherwise. There is room for further research to explore the benefits and challenges of developing bimodal queue management applications where the target queue length (or NVDZ) is variable.

The existence of the relationship between TMS and TCTTr has been confirmed in this thesis. The functional relationship between TMS and TCTTr can be quantified through regional transportation planning models, but a simpler, more practical approach is required in order to be able to quantify the impacts of bimodal management measures on the TMS.

The fuzzy logic based control algorithms for calculating the duration of green signal of the metering stage is promising. The frequent criticism of fuzzy logic control is the

absence of a more formal tuning process. In this thesis, a trial-and-error tuning process was implemented. Although, the tuning process was not impractically too complicated or lengthy, there is a general preference for more formalized methods. Research in this direction would be helpful.

The application of a fuzzy logic using ANFIS for tuning membership functions has been demonstrated in the thesis. However, further refinements of the methodology and tests would be useful. For example, one topic could be how to remove the bias that appeared in some of the tests. Another topic could further investigate the possibility of designing the data collection system and the tuning process so that the same membership function parameters could be used for different signal timings.

Although significant efforts have been invested in proving the queue management concept by demonstrating its elements through model and real-world examples, further testing and demonstration on different real-world examples would be useful. These examples could include intersections with more than one congested approach, longer congested corridors, or larger congested areas. Expanding the concept to these examples should not be too complex. However, testing and evaluating the traffic management techniques on a large congested network is not a trivial task. The difficulty of isolating and measuring the impacts of congestion and the effectiveness of control concepts and techniques increases exponentially with the size of the analysed network. The technical methods and principles develop in this thesis provide the required foundation for such a task.

12. REFERENCES

Abdulhai, B., Shalaby, A., Lee, J., Georgi, A. (2002), Microsimulation Modelling and Impact Assessment of Streetcar Transit Priority Options, The Toronto Experience, Transportation Research Board, Washington D.C.

Abu-Lebdeh, G., Benekohal, F. Rahim (1997) Development of Traffic Control and Queue Management Procedures for Oversaturated Arterials, Transportation Research Record 1603, Transportation Research Board, Washington D.C.

Abu-Lebdeh, G., Benekohal, F. Rahim (2000) Genetic Algorithms for Traffic Signal Control and Queue Management of Oversaturated Two-Way Arterials, Transportation Research Record 1727, Transportation Research Board, Washington D.C.

Akcelik, R., (1988) The Highway Capacity Manual Delay Formula for Signalized Intersections, ITE Journal, March

Akcelik, R. (1992) SIDRA User Guide, Report No. WD TE91/012B, Australian Road Research Board

Akcelik, R. (1993) Traffic Signals: Capacity and Timing Analysis, Research Report APR 123, Australian Road Research Board, Fifth Reprint

Akcelik & Associates (2007) Introducing SIDRA INTERSECTION 3,
http://www.sidrasolutions.com/Documents/Introducing_SIDRAIntersection.pdf (October 19, 2007)

Alberta Infrastructure and Transportation (2005) Traffic Impact Assessment Guidelines,
Transportation and Civil Engineering Division, Technical Standards Branch

Allsop, R.E. (1972) Delay at a Fixed Time Traffic Signal, Theoretical Analysis,
Transportation Science, Vol.6, pp. 260-285

Al-Mudhaffar, A., (2006) Impacts of Traffic Signal Control Strategies, Doctoral Thesis,
Royal Institute of Technology, Stockholm, Sweden

Andrade, K., Uchida, K., Kagaya, S. (2006) Development of Transport Mode Choice
Model by Using Adaptive Neuro-Fuzzy Inference System, Transportation Research
Record No 1977, Washington D.C.

Aristotle (1952) Metaphysics, Translated by Richard Hope, University of Michigan Press

Automobile Club of Michigan (1947) "Mr. Trafficlight", Motor News, March 1947 (pp.
14-15, 27) <http://www33.brinkster.com/iiii/trfclt/mrtrafficlight/mrtrafficlight.html>
(October 12, 2007)

BBC News, (2003) Huge Road Schemes for England, Wednesday, 9 July, 2003,
http://news.bbc.co.uk/2/hi/uk_news/3056636.stm (September 5 2008)

Beauchamp-Báez, G., Rodríguez-Morales, E., Muniz-Marrero, E. (1997) A Fuzzy Logic Based Phase Controller for Traffic Control, 0-7803-3796-4/97, IEEE

Ben-Akiva, M. and Lerman, S.R. (1985) Discrete Choice Analysis, Theory and Application to Predict Travel Demand, Cambridge, MA, MIT Press

Birst, S., Baker, J., Shouman, K. (2007) Comparison of Traffic Simulation Models to HCM 2000 Using Various Traffic Levels under Pre-timed Signal Control, Transportation Research Board, Washington D.C.

Bishop, R. (2001) "Whatever Happened to Automated Highway Systems (AHS)?", Traffic Technology International, August-September 2001, <http://faculty.washington.edu/jbs/itrans/bishopahs.htm> (October 12, 2007)

Black, J., Wanat, J. (1998) Traffic Signals and Control, Ramp Metering, and Lane Control Systems http://www.path.berkeley.edu/~leap/TTM/Traffic_Control/images/logosmall.gif (September 29, 2000)

Brampton, City of (2003) Six Pillars, Supporting Our Great City, Strategic Plan www.brampton.ca/headlines/six_pillars.pdf (September 17, 2008)

Brule F. James (1985) Fuzzy Systems – A Tutorial, <http://austinlinks.com/Fuzzy/tutorial.html>, (February 11, 2004)

Buchanan, C. (1963) Traffic in Towns, HMSO, London

Campbell, S., (2002) Zeno's Paradox of Plurality and Proof by Contradiction, Department of Education, University of California, Irvin,

<http://www.aug.edu/dvskel/Campbell2002.htm> (September 10, 2008)

Calgary, City of (2005) Calgary Transportation Plan 2005,

<http://content.calgary.ca/CCA/City+Hall/Business+Units/Transportation+Planning/Calgary+Transportation+Plan+2005/Calgary+Transportation+Plan+2005.htm> (September 5, 2008)

CITE (2008) Canadian Capacity Guide for Signalized Intersections, Third Edition, Institute of Transportation Engineers, District 7 - Canada

Clegg, R.G. and Clune, A.J (1999) MUSIC Project, Urban Traffic Control for Traffic Demand Management, Transportation Research Record 1682, Transportation Research Board, Washington D.C. <http://gridlock.york.ac.uk/music/docs/TRB.doc>

Cleveland Automobile Club (1914) The Motorist, New Traffic Signal Installed, August 1914, <http://www33.brinkster.com/iiii/trfclt/105th-euclid/105th-euclid.html> (October 12, 2007)

CUTR (2007) Best Practices for Traffic Incident Management in Florida, Centre for Urban Transportation Research, University of South Florida, <http://www.iacptechnology.org/IncidentManagement/BestPracticesFLDOT.pdf> (October 12, 2007)

Crochat, P., Franklin, D (2003) Back-Propagation Neural Network Tutorial,
http://iee.uow.edu.au/~daniel/software/libneural/BPN_tutorial/BPN_English/BPN_English/ (July 23, 2008)

Dion, F., Rakha, H., (2004) Integration of Transit Signal Priority within Adaptive Traffic Signal Control Systems, Transportation Research Board, Washington, D.C.

Duke, J., Schofer, J., May, A., (1967) A Statistical Analysis of Speed-Density Hypothesis, Highway Research Record 154, Transportation Research Board, Washington, D.C.

Edie, L. (1961) Car-Following and Steady state Theory for Non-Congested Traffic, Operations Research, Vol 9, ORSA, Washington, D.C.

Ellis, R. (1964) Analysis of Linear Relationships in Speed-Density and Speed-Occupancy Curves, Report, North Western University, Evanston, IL, December

Engelbrecht, R.J., Fambro, D.B., Roupail, N.M., Barkawi A.A. (1997) Validation of Generalized Delay Model for Oversaturated Conditions, Transportation Research Record, 1572, Transportation Research Board, Washington D.C.

Fambro, B., Roupail, N.M. (1997) Generalized Delay Model for Signalized Intersections and Arterial Streets, Transportation Research Record 1572, Transportation Research Board, Washington D.C.

FHWA (1991) Transyt-7F Users Guide, Methodology for Optimizing Signal Timing,
U.S. Department of Transportation, Federal Highway Administration

FHWA (1995) Improving Traffic Signal Operations- A Primer, US Department of
Transportation, Federal Highway Administration

FHWA (2001) Regional Traffic Incident Management Program, Implementation Guide,
Federal Highway Administration, May

http://www.itsdocs.fhwa.dot.gov/JPODOCS/REPTS_TE/13149.pdf (October 12, 2007)

Florida DOT (1997) Site Impact Handbook, April

Fraser, N. (1998) The Biological Neuron,

<http://www.virtualventures.ca/~neil/neural/neuron-a.html> (July 23, 2008)

Gal-Tzur, A., Mahalel, D., Prashker, Joseph N. (1993) Signal Design for Congested
Networks Based on Metering, Transportation Research Record 1398, Transportation
Research Board, Washington D.C.

Gazis, D.C. (1964) "Optimum Control of a System of Oversaturated Intersections",
Operations Research, No 12

Gazis, D.C., R.B. Potts (1965) "The Oversaturated Intersection", Proc. 2nd International
Symposium on the Theory of Road Traffic Flow, pp. 221-237, OECD, Paris

GESAMP (1986) Environmental Capacity, An Approach to Marine Pollution Prevention, Food and Agriculture Organization of the United Nations, Rome,

<http://www.fao.org/docrep/meeting/003/s0645e/s0645e00.htm> (September 17, 2008)

Girianna, M., Benekohal, R.F. (2004) Using Genetic Algorithms to Design Signal Coordination for Oversaturated Networks, Intelligent Transportation Systems, 8,117-129

Goodwin, P.B. (1997) "Solving Congestion", Inaugural Lecture for the Professorship of Transport Policy , University College London, <http://www.cts.ucl.ac.uk/tsu/pbginau.htm> (September 10, 2005)

Greenshields, G.(1998) A Study of Traffic Capacity, Proceedings of the Highway Research Board, Vol 14, Transportation Research Board, Washington D.C.

Hagen, T.L., Courage, G.C. (1989) Comparison of Macroscopic Models for Signalized Intersection Analysis, Transportation Research Record, 1225, Transportation Research Board, Washington D.C.

Hounsell, N.B., Kumara, S.S.P. (2005) Queue Re-location Using Pre-signals for Bus Priority: An Analytical Approach to Evaluate Isolated Pre-signals, Transportation Research Board, Washington, D.C.

HRB, (1965) Highway Capacity Manual, Special Report 87, Highway Research Board, Washington, D.C.

Huddart, K.W., C.C.Wright (1989) Catastrophic traffic congestion and some possible ways of preventing it, Proc., Traffex Seminar on Congestion, control and parking enforcement, PTRC Education and Research Services, Ltd, London

Hurdle, V.F. (1984) Signalized Intersection Delay Models-A Primer for the Uninitiated, Transportation Research Record 971, Transportation Research Board, Washington, D.C.

ITE (2006) Transportation Impact Analyses for Site Development: An ITE Proposed Recommended Practice, Institute of Transportation Engineers

Jamshidi,M., Vadiie T, Ross,T.J. (1993) Fuzzy Logic and Control, Prentice Hall

Jang, J.S-R. ANFIS (1993) Adaptive-Network-Based Fuzzy Inference Systems, IEEE Transactions on Systems, Man, and Cybernetics, Vol 23, No. 3, May 1993, pp. 665-685

Joos, E. (Undated) Economy and Ecology are no Contradictions, Three messages from Zurich concerning the new transport policy, Zurich Transport Authority

Jowit, J. (2007) M1 widening to cost £21m per mile, The Observer,
<http://www.guardian.co.uk/uk/2007/may/06/transport.world>, (September 5, 2008)

Khatib, Z.K., Kyte, M. (2001) Uncertainty in projecting the level of Service of Signalized and Unsignalized Intersections, Transportation Research Board, Washington D.C.

Kikuchi, S., Miljkovic, D. (1999) Method To Process Observed Traffic Data for Consistency, Application of Fuzzy Optimization Concept, Transportation Research Record, 1679, Transportation Research Board, Washington D.C.

Kikuchi, S., Miljkovic, D., Zuylen, H.J. (2000) Examination of Methods That Adjust Observed Traffic Volumes on a Network, Transportation Research Record, 1717, Transportation Research Board, Washington D.C.

Kittelsohn, W., K., Roess, R.P. (2001) Highway Capacity Analysis After the HCM 2000, Transportation Research Board, Washington D.C.

Kittelsohn, W.K.(2000) Historical Overview of the Committee on Highway Capacity and Quality of Service, Transportation Research Circular E-C018: 4th International Symposium on Highway Capacity, Fourth International Symposium on Highway Capacity, Maui, Hawaii, Transportation Research Board, June 27–July 1
http://onlinepubs.trb.org/onlinepubs/circulars/ec018/01_63.pdf (October 10, 2007)

Kosko, B. (1993) Fuzzy Thinking, The New Science of Fuzzy Logic, Hyperion, New York

Kosonen, I. (2003) Multi-agent Fuzzy Signal Control Based on Real-time Simulation, Transportation Research Part C 11, 389-403

KPMG Peat Marwick (1992) Compendium of Travel Demand Forecasting Methodologies, Federal Transit Administration, Washington, DC

Kumara, S.S.P., Hounsell, N.B., Cherrett, T.J. (2007) Evaluating Bus Priority and Queue Relocation Techniques of Pre-signals, Transportation Research Board, Washington, D.C.

Lavine, T. Z. (1984) "Shadow and Substance - Plato's Sources: The Pre-Socratics", From Socrates to Satire: The Philosophic Quest. New York, New York, Bantam Books, 24.

Lawrence T. H., Kenneth G. C. (1989) Comparison of Macroscopic Models for Signalized Intersection Analysis, Transportation Research Record 1225, Transportation Research Board, Washington, D.C.

Lejewski, C. (1967) "Jan Lukasiewicz", Encyclopedia of Philosophy, Vol. 5, MacMillan, NY, pp 104-107

Lewis, H. W. (1997) The Foundations of Fuzzy Control, Plenum Press, New York

Lewis, P. H., Yang, C. (1997) Basic Control Systems Engineering, Prentice-Hall, Inc.

Lieberman, B. Edwards, C, Jinil, and Prassess, Shenk E. (2000) Formulation of Real-Time Control Policy for Oversaturated Arterials, Transportation Research Record 1727, Transportation Research Board, Washington, D.C.

Longley, D. (1968) A Control strategy for computer-controlled traffic networks, Transportation Research 2, 391-408

Mamdani, E. Assilian, S. (1975) "An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller", International Journal of Man-Machine Studies, Vol 7, 1-13

MathWorks (2001) Fuzzy Logic Toolbox, User's Guide Versions2

May, A. D., Montgomery, F.O., Quinn,D.J. (1998) "Control of congestion in highly congested networks", Proc. CODATU IV Conference, Jakarta, June

May, A.D. (1991) "Queue management project: Strategies for the management of queues at upstream junctions", ITS working paper WP346, Institute of Transport Studies, University of Leeds

Meyer, M.D., Miller, J.E. (2001) Urban Transportation Planning, A decision-Oriented Approach, Second Edition, McGraw-Hill series in transportation

Motorist, The (1914) New Traffic Signal Installed, Cleveland Automobile Club,
<http://www33.brinkster.com/iiii/trfclt/105th-euclid/105th-euclid.html> cited at
<http://www33.brinkster.com/iiii/trfclt/> (September 12, 2008)

Motor News (1947) Mr. 'Trafficlight', Automobile Club of Michigan, March,
<http://www33.brinkster.com/iiii/trfclt/mrtrafficlight/mrtrafficlight.html> cited at
<http://www33.brinkster.com/iiii/trfclt/> (September 12, 2008)

Mucsi, K., Khan, A (2002) Diminishing Marginal Capacity Benefits of Additional Lanes, Transportation Research Board, Washington D.C.

Mucsi, K. (2004) Evaluation of Transit Priority Measures in the Woodroffe Corridor, Unpublished internal report, City of Ottawa

Mucsi, K. (2005) Transit Priority Background, Unpublished internal report, City of Ottawa

Mucsi, K. (2008) The Tale of Seven Queue Jumps, ITE Annual Meeting, Anaheim, California

Mueller, E.A. (1970) Aspects of the History of Traffic Signals, IEEE Transactions on Vehicular Technology, Vol. VT 19, No. 1

Millard, F (2008) Reigning Supreme, How to Rule at the Intersection, Traffic Technology International, February/March

NATO (1976) Committee on the Challenges of Modern Society, Bus Priority Systems, CCMS Report 45, Transportation and Road Research Laboratory, Department of the Environment, United Kingdom

NCHRP (1992) Internal Metering Policy for Oversaturated Networks, Volume 1 and Volume 2, NCHRP 03-38(4) Unpublished Report, Transportation Research Board, Washington, D.C.

NCHRP (1998) Travel Estimation Techniques for Urban Planning, NCHRP Report 365, Transportation Research Board, Washington, D.C.

NCHRP (2007) Operation of Traffic Signal Systems in Oversaturated Conditions – Future project, NCHRP 03-90, 2007 (Pending, October 18, 2007) Transportation Research Board, Washington, D.C.

Neuron (2008) <http://en.wikipedia.org/wiki/Neuron> (July 23, 2008)

Newell, G. (1989) Theory of Highway Traffic Signals, University of California at Berkeley

New York Times (February 6, 1920) “5th Av. Traffic to Go South From 10 to 5, Dr. Harriss Explains Details of “One-Way” Experiment to Begin Feb 16”, p.26

<http://www33.brinkster.com/iiii/trfclt/NYT-feb-06-1920.html> cited at

<http://www33.brinkster.com/iiii/trfclt/> (October 12, 2007)

New York Times (March 15, 1920) “The Traffic Lights”

<http://www33.brinkster.com/iiii/trfclt/NYT-mar-15-1920.html> cited at

<http://www33.brinkster.com/iiii/trfclt/> (October 12, 2007)

New York Times (January 5, 1922) “To Rule All Traffic From Times Square, Dr. Harriss Plans to Push Button Stop Every Auto From 14th St. to Harlem”

<http://www33.brinkster.com/iiii/trfclt/NYT-jan-05-1922.html>, cited at

<http://www33.brinkster.com/iiii/trfclt/> (October 12, 2007)

Niittymäki, J, Kikuchi, S. (1998) Application of Fuzzy Logic to the Control of a Pedestrian Crossing Signal, Transportation Research Record 1651, Transportation Research Board, Washington, D.C.

Niittymäki, J. (2001) General Fuzzy Rule Base for Isolated Traffic Signal Control – Rule Formulation, Transportation Planning and Technology, Vol 24, pp. 227-247

Niittymäki, J., Kosonen, T., Nevala, R. (2001) Fuzzy Traffic Signal Control in Major Arterials, IEEE Intelligent Transportation Systems Conference Proceedings, Oakland, Ca, August 25-29

Oakes, J., Metzger, D. (1995) “Park View Pre-signals in Uxbridge Road, Ealing”, Traffic Engineering +Control, February

Ottawa, City of (1997) Transportation Master Plan, Ottawa

Ottawa, City of (2003a) Transportation Master Plan, Ottawa

Ottawa, City of (2003b) City of Ottawa Official Plan, As adopted by City Council May 2003

Ottawa, City of (2006) Transportation Impact Assessment Guidelines, Department of Public Works and Services, Department of Planning and Growth Management, October, Ottawa

Owens, S. (1995) From 'predict and provide' to 'predict and prevent'?, Pricing and planning in transportation policy, Transport Policy, Vol 2, No.1.pp43-49

Pant, S.N., Holbert, K.E. (2007) Fuzzy Logic in Decision Making and Signal Processing, Arizona State University, Fulton School of Engineering,
<http://enpub.fulton.asu.edu/powerzone/FuzzyLogic/index.htm> (August 25, 2007)

Park, B. (2002) Hybrid Neuro-Fuzzy Application in Short-Term Freeway Traffic Volume Forecasting, Transportation Research Record 1802, Washington D.C.

Parsons Brinckerhoff Quade & Douglas (1992) Inc., Review of Best Practices, Washington, DC

Papageorgiou, M., Hadj Salem, H., Middelham, F., (1997) ALINEA Local Ramp Metering, Summary of Field Results, Transportation Research Record 1603, Transportation Research Board, Washington D.C.

Papageorgiou, M., Vigos, G. (2008) Relating time –occupancy measurements to space-occupancy and link vehicle-count, Transportation Research, Part C, 16 1-17

Pappis, C., Mamdani, E (1977) A Fuzzy Logic Controller for a Traffic Junction, IEEE Transactions and Systems, Man and Cybernetics, ASMC 7, 707-717

Passiono, K.M., Yurkovich, S. (1998) Fuzzy Control, Addison Wesley Longman Inc.

Pavis, A., Saracena, A., Yagoda, H.N., Bauer, A. (1994) Congestion Management Through Bus Metering at the Lincoln Tunnel, Transportation Research Record 1496, Transportation Research Board, Washington, D.C.

Pearson, R. (2003) Intelligent Transportation Systems – Traffic Signal Control, UC Berkeley Institute of Transportation Studies and the California Department of Transportation,
http://www.calccit.org/itsdecision/serv_and_tech/Traffic_signal_control/traffsigrep_print.htm (October 13, 2007)

Pearson R, Black J., and Wanat, J. (2003) Ramp Metering, California Center for Innovative Transportation, UC Berkeley Institute of Transportation Studies and the California Department of Transportation
http://www.calccit.org/itsdecision/serv_and_tech/Ramp_metering/ramp_metering_report.htm#r_meter_rates, (October 12, 2007)

Peel, Region of (2005) Guidelines for the Preparation of Transportation Impact Studies, January

Pignataro, L. J., W.R. McShane, K.W. Crowley (1978) Traffic Control in Oversaturated Street Networks, NCHRP Report 194, Transportation Research Board, Washington, D.C.

Pitsiava-Latinopoulou, M., Mustafa, A.A.S (1992) "The accuracy of estimating delays at signalised intersections: A comparison between two methods", Traffic Engineering + Control, May

Prassas, S.E. (2002) A Comparison of Delay on Arterials by Simulation and HCM Computations, Transportation Research Board, Washington D.C.

Přibyl, A., Goulias K.G. (2003) Application of Adaptive Neuro-Fuzzy Inference System to Analysis of Travel Behaviour, Transportation Research Record 1854, Transportation Research Board, Washington D.C.

Robertson, D.I. (1979) Traffic Models and Optimum Strategies of Control – A Review, Proceedings, International Symposium on Traffic Control System, Edited by Homburger, W.S., Steinman, L. Vol 1, Berkeley, California, pp. 262-288.

Rydin, Y. (1998) Land Use Planning and Environmental Capacity: Reassessing the Use of Regulatory Policy Tools to Achieve Sustainable Development, Journal of Environmental Planning and Management, Volume 41, Issue 6, 749-765

Quinn, D.J. (1992) Review of Queue Management Strategies, Traffic Engineering+ Control, November

Roess, R.P., McShane, R.W., Prassas, E.S. (1998) Traffic Engineering, Second Edition, Prentice Hall

Rouphail, N.M., Akcelik, R.(1996) "More Signal Capacity with Less Green?" Institute of Transportation Engineers Journal, July

Sessions, G., (1971)Traffic Devices, Historical Aspects Thereof, Washington DC: Institute of Traffic Engineers, c. <http://www33.brinkster.com/iiii/trfcl/> (October 12, 2007)

SCOOT (2007) <http://www.scoot-utc.com/> (July 15, 2007)

Schmidt, Robert E. and M. Earl Campbell, (1956) Highway Traffic Estimation, Eno Foundation for Highway Traffic Control, Saugatuck, CT.

Smeed, R., Wardrop, A. (1964) An exploratory comparison of the relative advantage of cars and buses in urban areas, Institute of Transport Journal 30(9) pg 301-305

Stanford Encyclopedia of Philosophy (2007) Prisoner's Dilemma, <http://plato.stanford.edu/entries/prisoner-dilemma/> (September 2, 2008)

Stufflebeam Row, K., LaDow, E., and Moler, S. (2004) Glenwood Canyon 12 Years Later, Public Roads, March/April 2004, <http://www.tfhr.gov/pubrds/04mar/04.htm>, (September 5, 2008)

Spiegel, M.(1972) Theory and Problems of Statistics, Schaum's Outlines Series, McGraw-Hill Book Company

Takagi, T., Sugeno, M.(1985) Fuzzy Identification of Systems and its Applications to Modeling and Control, IEEE Transactions on Systems, Man and Cybernetics 15(1)16-132

Tarko, A.P. (2000) Analyzing Arterial Streets in Near-Capacity or Overflow Conditions, Transportation Research Record 1710, Transportation Research Board, Washington D.C.

Taylor C., Meldrum, D., Jacobson, L. (1998) Fuzzy Ramp Metering, Design Overview and Simulation Results, Transportation Research Record 1634, Transportation Research Board, Washington, D.C.

Teodorovic, D., Vukadinovic, K.(1998) Traffic Control and Transport Planning: A Fuzzy Sets and Neural Networks Approach, Kluwer Academic Publishers

Tian Z.Z., Urbanik II, T., Engelbrecht, R., Balke K.(2002) Variations in Capacity and Delay Estimates from Microscopic Traffic Simulation Models, Transportation Research Record 1802, Transportation Research Board, Washington D.C.

Toronto, City of (2003) Guidelines for the Preparation of Impact Studies, Toronto

Toronto, City of (2007) Toronto Official Plan,

http://www.toronto.ca/planning/official_plan/introduction.htm (September 5, 2008)

Trafficware (2007) "Queue Interactions in Synchro 6",

http://www.trafficware.com/assets/pdfs/Queue_Interactions_Document2.doc , (August 7, 2007)

TRB (1985) Highway Capacity Manual, Special Report 209, Transportation Research Board, Washington D.C.

TRB (2000) Highway Capacity Manual, Transportation Research Board, Washington D.C.

TRB (2007) Metropolitan Travel Forecasting, Current Practice and Future direction, Special Report 288, Transportation Research Board, Washington D.C.

<http://onlinepubs.trb.org/onlinepubs/sr/sr288.pdf> (February 24, 2008)

TRC (1991) Transyt-7F, Users Guide, Methodology for Optimizing Signal Timing, MOST, Volume 4, University of Florida, Transportation Research Center

TTI (1991) Passer-II-90 Program User's Guide, Texas Transportation Institute, College Station, Texas

Underwood, R., (1961) Speed, Volume, and Density Relationships, Quality and Theory of Traffic Flow, Yale Bureau of Highway Traffic Flow, Yale Bureau of Highway Traffic, New Haven, CT

UK DOT (1997) Keeping Buses Moving, A guide to traffic management to assist buses in urban areas, Local Transport Note, 1/97, United Kingdom, Department of Environment, London

UK DOT (1999) Traffic: The “SCOOT” Urban Traffic Control System, Advisory Leaflets, 7/99, UK Department of Transport,
http://www.roads.dft.gov.uk/roadnetwork/ditm/tal/traffic/07_99/index.htm (September 16, 2002)

Upham, P. (2001) Environmental Capacity of Aviation: Theoretical Issues and Basic Research Directions, *Journal of Environmental Planning and Management*, Volume 44, Issue 5, 721-734

Urbanik, T. (2007) Oversaturation: What do We Really Know?, Presentation at Transportation Research Board, Washington D.C.

Urban Transportation Monitor (2003) Survey of Ramp Metering Systems

Vancouver, City of (2007) Transportation Plan Update: A Decade of Progress,
<http://vancouver.ca/engsvcs/transport/plan/index.htm> (September 5, 2008)

Vigos, G., Papageorgiou, M., Wang, Y., (2008) Real-time estimation of vehicle-count within signalized links, *Transportation Research, Part C*, 16 18-35

VISSIM 5.0 (2007) User Manual, PTV

Voorhees, A. M., (1956) A General Theory of Traffic Movement, *Proceedings, Institute of Traffic Engineers*, New Haven, CT.

Washburn, S., Larson, N. (2001) Signalized Intersection Delay Estimation: Case Study Comparison of Transyt-7F, Synchro, and HCS, Transportation Research Board, Washington, D.C.

Weiner, E. (1992) Urban Transportation Planning in the United States: An Historical Overview, A Revised Edition, Praeger/Greenwood

Webster, F.V. (1958) Traffic Signal Settings, Road Research Technical Report 39, Road Research Laboratory, Her Majesty Stationary Office, London, England

Webster, F.V. Cobbe, B.M.,(1966) Traffic Signals, Road Research Technical Report 56, Road Research Laboratory, Her Majesty Stationary Office, London, England

Yen, J., Langari, R.(1999) Fuzzy Logic, Intelligence, Control, and Information, Prentivce Hall, ISBN 0-13-525817-0 (pg. 68)

Zadeh, L.A. (1965) "Fuzzy Sets", Information and Control, Vol 8, 338-353

Zhang, L., Li, H., Prevedouros, P.D. (2005) Signal Control for Oversaturated Intersections Using Fuzzy Logic, 05-0156 Transportation Research Board, Washington, D.C.

13. APPENDIXES

Appendix A: Delay Calculation with the Deterministic Queuing Model

Table 13.1: Type A Arrival Distribution

Hours	1	Red	Green	Green
Cycle:	60 sec	Start	Start	End
Green:	30 sec/cycle	0	30	60
Demand:	700 pcu / 1 hr			
Measured SF:	1800 pcu/hr (all lanes)			
Capacity:	900 pcu/hr (all lanes)	75.0 veh/period	TotalDelay: 2.39 veh-hr	
D/C:	0.778 1 hr			AverageDelay: 12 sec/veh

Historic Volume Distribution					100.00%	
Period:	0:05 min====> 300		Distrib.	Projection		
	Interval	2003 Vol		vol	D/C	
7:00	7:05	1	50	8.33%	58	0.78
7:05	7:10	2	50	8.33%	58	0.78
7:10	7:15	3	50	8.33%	58	0.78
7:15	7:20	4	50	8.33%	58	0.78
7:20	7:25	5	50	8.33%	58	0.78
7:25	7:30	6	50	8.33%	58	0.78
7:30	7:35	7	50	8.33%	58	0.78
7:35	7:40	8	50	8.33%	58	0.78
7:40	7:45	9	50	8.33%	58	0.78
7:45	7:50	10	50	8.33%	58	0.78
7:50	7:55	11	50	8.33%	58	0.78
7:55	8:00	12	50	8.33%	58	0.78
Total			600	100.00%	700	
Peak HR**			200		233	
Peak Period Factor:				PPF	1.00	

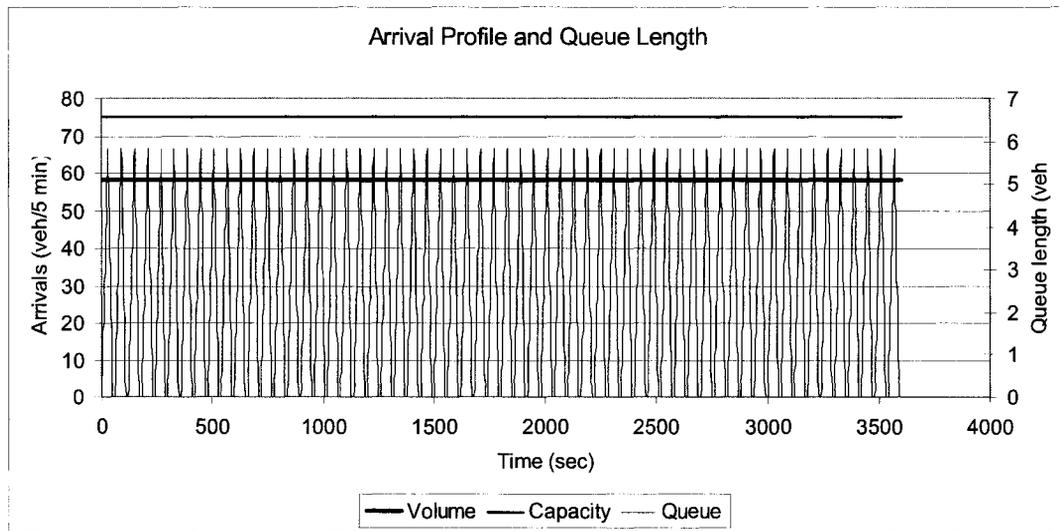


Figure 13.1: Type A Arrival Distribution: Arrivals and Queue Length

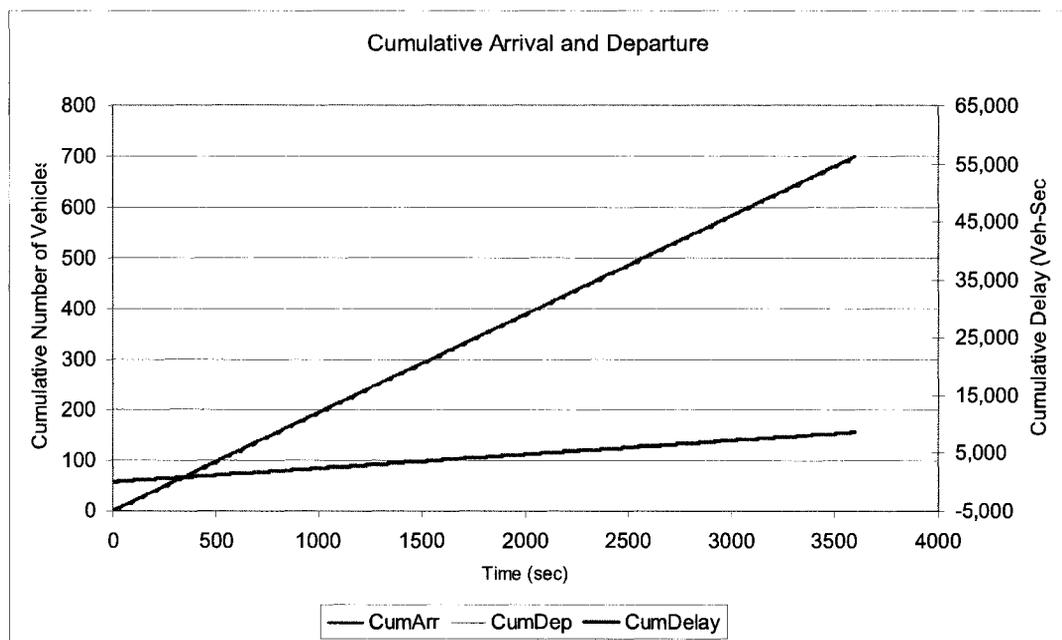


Figure 13.2: Type A Arrival Distribution: Cumulative Arrivals and Departures

Table 13.2: Type B Arrival Distribution

Hours	1	Red	Green	Green
Cycle:	60 sec	Start	Start	End
Green:	30 sec/cycle	0	30	60
Demand:	700 pcu / 1 hr			
Measured SF:	1800 pcu/hr (all lanes)			
Capacity:	900 pcu/hr (all lanes)	75.0 veh/period	TotalDelay: 13.22 veh-hr	
D/C:	0.778 1 hr			AverageDelay: 68 sec/veh

Historic Volume Distribution				100.00%		
Period:	0:05 min====>	300	Distrib.	Projection	vol	D/C
	Interval	2003 Vol				
7:00	7:05	1 50	4.17%		29	0.39
7:05	7:10	2 50	4.17%		29	0.39
7:10	7:15	3 100	8.33%		58	0.78
7:15	7:20	4 100	8.33%		58	0.78
7:20	7:25	5 100	8.33%		58	0.78
7:25	7:30	6 100	8.33%		58	0.78
7:30	7:35	7 150	12.50%		88	1.17
7:35	7:40	8 150	12.50%		88	1.17
7:40	7:45	9 150	12.50%		88	1.17
7:45	7:50	10 150	12.50%		88	1.17
7:50	7:55	11 50	4.17%		29	0.39
7:55	8:00	12 50	4.17%		29	0.39
Total		1,200	100.00%		700	
Peak HR**		600			350	
Peak Period Factor: PPF				1.50		

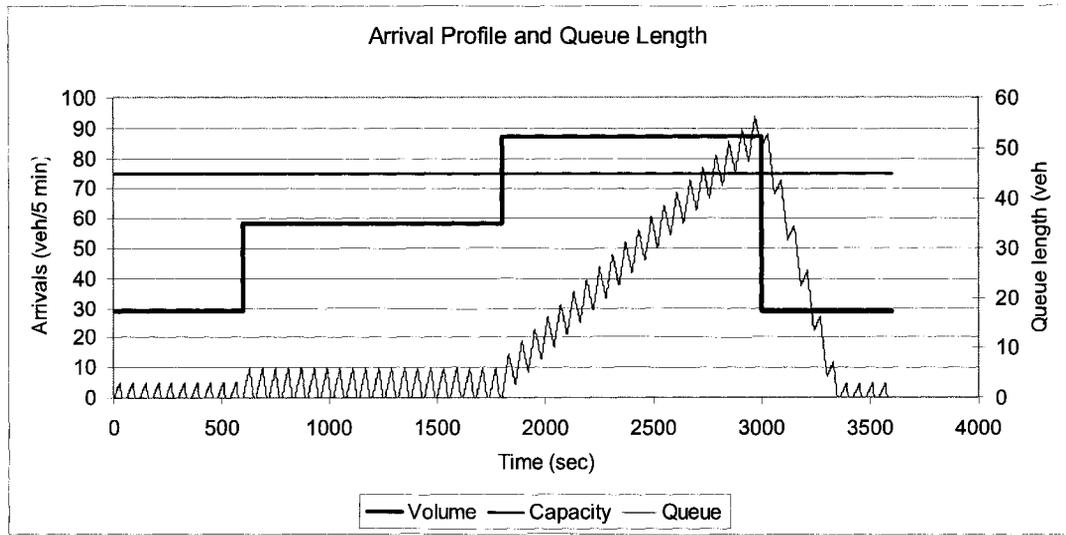


Figure 13.3: Type B Arrival Distribution: Arrivals and Queue Length

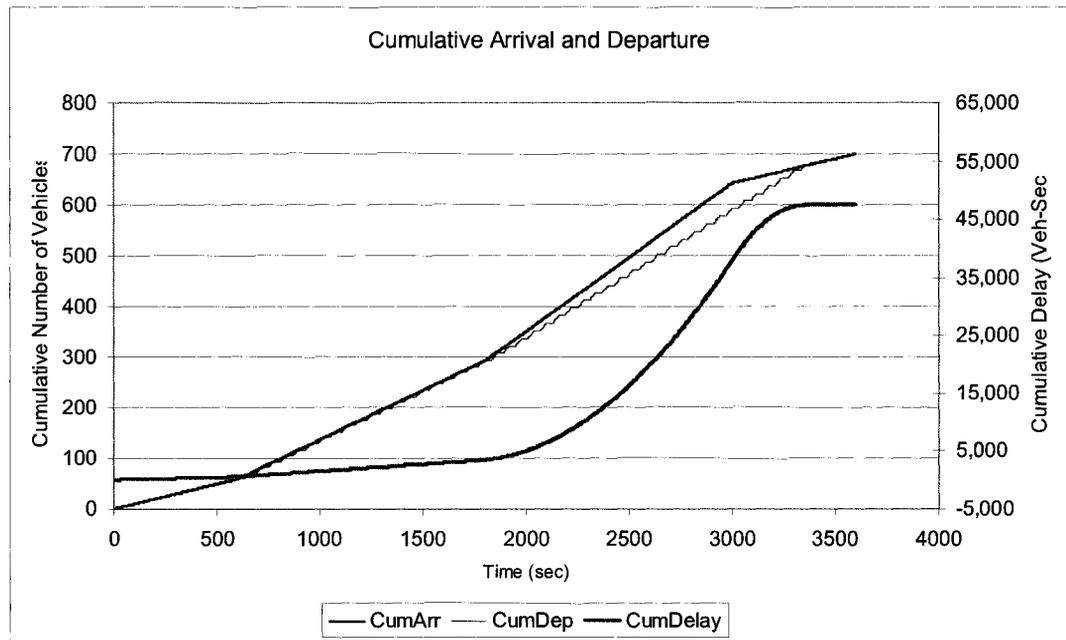


Figure 13.4: Type B Arrival Distribution: Cumulative Arrivals and Departures

Table 13.3: Type C Arrival Distribution

Hours:	1	Red	Green	Green
Cycle:	60 sec	Start	Start	End
Green:	30 sec/cycle	0	30	60
Demand:	700 pcu / 1 hr			
Measured SF:	1800 pcu/hr (all lanes)			
Capacity:	900 pcu/hr (all lanes)	75.0 veh/period	TotalDelay: 17.27 veh-hr	
D/C:	0.778 1 hr			AverageDelay: 89 sec/veh

Historic Volume Distribution					100.00%
Period:	0:05 min====>	300	Distrib.	Projection	
	Interval	2003 Vol		vol	D/C
7:00	7:05	1 50	4.17%	29	0.39
7:05	7:10	2 50	4.17%	29	0.39
7:10	7:15	3 150	12.50%	88	1.17
7:15	7:20	4 150	12.50%	88	1.17
7:20	7:25	5 150	12.50%	88	1.17
7:25	7:30	6 150	12.50%	88	1.17
7:30	7:35	7 100	8.33%	58	0.78
7:35	7:40	8 100	8.33%	58	0.78
7:40	7:45	9 100	8.33%	58	0.78
7:45	7:50	10 100	8.33%	58	0.78
7:50	7:55	11 50	4.17%	29	0.39
7:55	8:00	12 50	4.17%	29	0.39
Total		1,200	100.00%	700	
Peak HR**		600		350	
Peak Period Factor:			PPF	1.50	

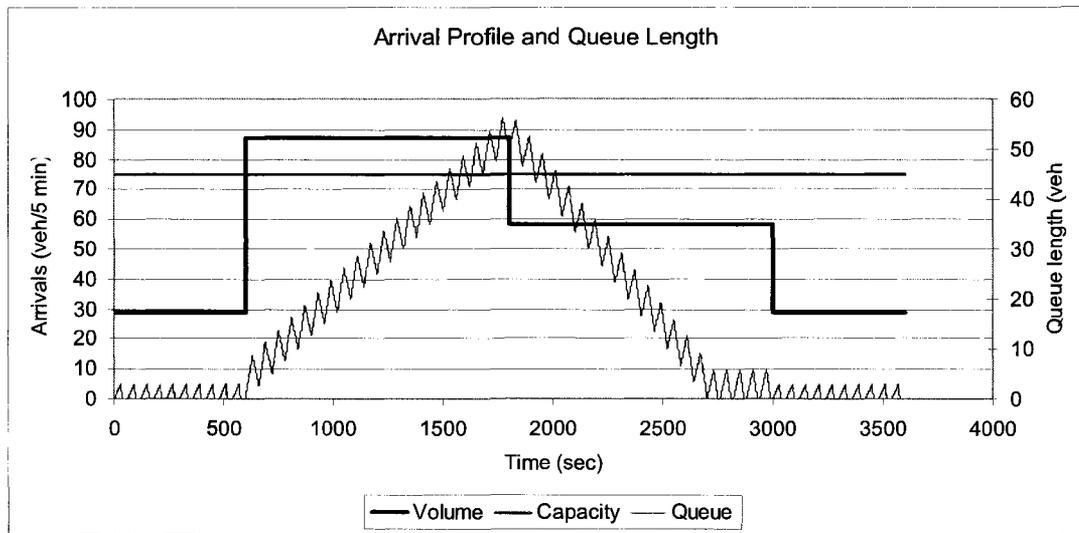


Figure 13.5: Type C Arrival Distribution: Arrivals and Queue Length

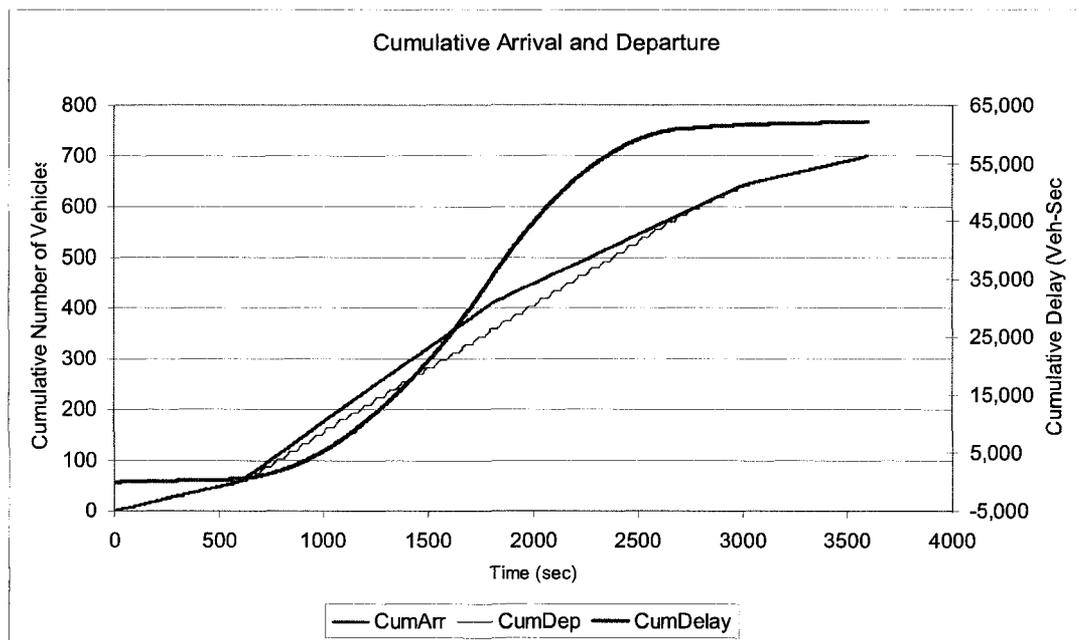


Figure 13.6: Type C Arrival Distribution: Cumulative Arrivals and Departures

Appendix B: Mean Delay Calculation with the Analytical Formulas

Table 13.4: Input Parameters

Input Parameters	Values
Demand (veh/hr)	700
Cycle (sec)	60
Green (sec)	30
Saturation flow rate (veh/hr)	1800
PHF	0.662

Webster's delay formula

$$d = d_1 + d_2 - d_3$$

$$d_1 = \frac{cyc \times \left[1 - \frac{g}{cyc} \right]^2}{2 \times \left[1 - \frac{g}{cyc} \times X \right]}$$

$$d_2 = \frac{X^2}{2 \times q \times (1 - X)}$$

$$d_3 = 0.65 \times \left(\frac{cyc}{q^2} \right)^{\frac{1}{3}} \times X^{2 + 5 \times \frac{g}{cyc}}$$

$$C = s \times \frac{g}{cyc}$$

$$X = \frac{V}{C}$$

$$q = \frac{V}{3600}$$

Calculations:

$$q = \frac{700}{3600} = 0.194 \text{ veh/sec}$$

$$C = 1800 \times \frac{30}{60} = 900 \text{ veh/hr}$$

$$X = \frac{700}{900} = 0.778$$

$$d_1 = \frac{60 \times \left[1 - \frac{30}{60}\right]^2}{2 \times \left[1 - \frac{30}{60} \times 0.778\right]} = 12.27 \text{ sec/veh}$$

$$d_2 = \frac{0.778^2}{2 \times 0.194 \times (1 - 0.778)} = 7.03 \text{ sec/veh}$$

$$d_3 = 0.65 \times \left(\frac{30}{0.194^2}\right)^{\frac{1}{3}} \times 0.778^{2 + 5 \times \frac{30}{60}} = 1.95 \text{ sec/veh}$$

$$d = 12.27 + 7.03 - 1.95 = 17.35 \text{ sec/veh}$$

Canadian Capacity Guide delay formula:

$$d = k_f \times d_1 + d_2$$

$$d_1 = \frac{\text{cyc} \times \left[1 - \frac{g}{\text{cyc}}\right]^2}{2 \times \left[1 - x_1 \times \frac{g}{\text{cyc}}\right]}$$

$$d_2 = 15 \times t_e \times \left[(X - 1) + \sqrt{(X - 1)^2 + \frac{240 \times X}{C \times t_e}} \right]$$

Calculations:

$$d_1 = \frac{60 \times \left[1 - \frac{30}{60}\right]^2}{2 \times \left[1 - \text{MIN}\left(1, \frac{700}{900}\right) \times \frac{30}{60}\right]} = 12.27 \text{ sec/veh}$$

$$d_2 = 15 \times 15 \times \left[(0.778 - 1) + \sqrt{(0.778 - 1)^2 + \frac{240 \times 0.778}{900 \times 15}} \right] = 6.57 \text{ sec/veh}$$

$$d = k_f \times d_1 + d_2 = 1 \times 12.27 + 6.57 = 18.84$$

k_f – adjustment factor for progression (=1 for isolated intersection)

x_1 – minimum of (1.0, V/C)

HCM 2000 formula:

$$d = PF \times d_1 + d_2 + d_3$$

$$d_1 = \frac{0.5 \times \text{cyc} \times \left[1 - \frac{g}{\text{cyc}}\right]^2}{1 - \text{MIN}\left(1, X\right) \times \frac{g}{\text{cyc}}}$$

$$d_2 = 900 \times t_e \times \left[(X - 1) + \sqrt{(X - 1)^2 + \frac{8 \times k \times l \times X}{C \times t_e}} \right]$$

PF – adjustment factor for progression (=1 for isolated intersection)

k – incremental delay adjustment for actuated control ($k=0.5$ for fixed time control)

l – incremental delay adjustment for the filtering or metering by upstream signals ($l=1$ for isolated intersections)

d_3 – initial queue delay at start of analysis period

Calculations:

$$d_1 = \frac{0.5 \times 60 \times \left[1 - \frac{30}{60}\right]^2}{1 - 0.778 \times \frac{30}{60}} = 12.275$$

$$d_2 = 900 \times 0.25 \times \left[(0.778 - 1) + \sqrt{(0.778 - 1)^2 + \frac{8 \times 0.5 \times 1 \times 0.778}{900 \times 0.25}} \right] = 6.57$$

$$d_3 = 0$$

$$d = 1 \times 12.28 + 6.57 + 0 = 18.8$$

Alternative HCM Delay Calculation Approaches

Approach 1: Based on the traditional method by using the 15 min peak period volume

$$V_{\text{peak}} = V / \text{PHF} = 700 / 0.662 = 1057 \text{ veh/hr}$$

$$t_a = 0.25 \text{ hr}$$

$$X = \frac{1057}{900} = 1.174$$

$$d_1 = \frac{60 \times \left[1 - \frac{30}{60}\right]^2}{2 \times \left[1 - \text{MIN}\left(1, \frac{1057}{900}\right) \times \frac{30}{60}\right]} = 15 \text{ sec/veh}$$

$$d_2 = 900 \times 0.25 \times \left[(1.174 - 1) + \sqrt{(1.174 - 1)^2 + \frac{8 \times 0.5 \times 1 \times 1.174}{900 \times 0.25}} \right] = 90.04 \text{ sec/veh}$$

$$d = 1 \times 12.27 + 6.89 = 105.04 \text{ sec/veh}$$

Approach 2: Based on the mean hourly volume and 60 min evaluation period

$$V=700 \text{ veh/hr}$$

$$t_a=1 \text{ hr}$$

$$d_1 = \frac{60 \times \left[1 - \frac{30}{60}\right]^2}{2 \times \left[1 - \text{MIN}\left(1, \frac{700}{900}\right) \times \frac{30}{60}\right]} = 12.27 \text{ sec/veh}$$

$$d_2 = 900 \times 1 \times \left[(0.778 - 1) + \sqrt{(0.778 - 1)^2 + \frac{8 \times 0.5 \times 1 \times 0.778}{900 \times 1}} \right] = 6.89 \text{ sec/veh}$$

$$d = 1 \times 12.27 + 6.89 = 19.16$$

Approach 3:

To facilitate the analysis of oversaturated conditions, the HCM 2000 suggests a procedure for the estimation of vehicle delay in oversaturated conditions. The Manual (TRB, 2000) suggests breaking down the analysis period that includes periods of oversaturation to shorter time periods with equal lengths that have relatively constant arrival flow rates. Thus, the analysis time, t_a becomes equal to the length of the period, and the arrival flow rate in the formula is the mean arrival flow rate for that period. The analysis is done for individual time periods. The HCM 2000 formula also has a delay component which calculates delays for the leftover queues from the previous period. With these modifications, the HCM 2000 provides an analytical procedure for the calculation of vehicle delay in oversaturated conditions. However, the procedure description in the Manual is incomplete.

Appendix C: Calculation of Transit Modal Share for different Transit Travel Time Changes

Table 13.5: Impact of Travel Time Changes on Transit Modal Share

IMPACT OF TRAVEL TIME SAVINGS ON TMS ASSUMING CHANGES TO IVTT ONLY																
Baseline		Change of utility					New Mode Splits									
Mode split		CAR IVTT		BUS IVTT			d(u)		Car		Bus		Car+Bus		IVTT	
Cars	Bus	d(IVTT)	Coef	d(u)	d(IVTT)	Coef	d(u)	e ^u	P ^{e^u}	e ^u	P ^{e^u}		P ^{car}	P ^{bus}		
0.90	0.10	0.0	-0.025	0	0.0	-0.025	0	1.00	0.90	1.00	0.100	1.000	0.900	0.100		
		0.0		0	-4		0.1	1.00	0.90	1.11	0.11	1.011	0.89	0.11		
		0.0		0	-6		0.15	1.00	0.90	1.16	0.12	1.016	0.89	0.11		
		0.0		0	-8		0.2	1.00	0.90	1.22	0.12	1.022	0.88	0.12		
		0.0		0	-10		0.25	1.00	0.90	1.28	0.13	1.028	0.88	0.125		
		0.0		0	-12		0.3	1.00	0.90	1.35	0.13	1.035	0.87	0.13		
		0.0		0	-14		0.35	1.00	0.90	1.42	0.14	1.042	0.86	0.14		
		0.0		0	-16		0.4	1.00	0.90	1.49	0.15	1.049	0.86	0.14		
		0.0		0	-18		0.45	1.00	0.90	1.57	0.16	1.057	0.85	0.15		
		0.0		0	-20		0.5	1.00	0.90	1.65	0.16	1.065	0.85	0.15		

IMPACT OF TRAVEL TIME SAVINGS ON TMS ASSUMING CHANGES TO IVTT AND OVTT																		
Number of buses		IVTT saving is in one direction only - same direction as car travel time is measured																
Orig roundtrip time		60 min																
Original wait time		30 min																
Original wait time		30 min																
Baseline		Change of utility					BUS OVTT			New Mode Splits								
Mode split		CAR IVTT		BUS IVTT			BUS OVTT		d(u)		Car		Bus		Car+Bus		IVTT +OVTT n=1	
Cars	Bus	d(IVTT)	Coef	d(u)	d(IVTT)	Coef	d(IVTT)	Coef	d(u)	e ^u	P ^{e^u}	e ^u	P ^{e^u}		P ^{car}	P ^{bus}		
0.90	0.10	0.0	-0.025	0	0.0	-0.025	0	-0.05	0	1.00	0.90	1.00	0.10	1.000	0.90	0.10		
		0.0		0	-4		-2		0.2	1.00	0.90	1.22	0.12	1.022	0.88	0.12		
		0.0		0	-6		-3		0.3	1.00	0.90	1.35	0.13	1.035	0.87	0.13		
		0.0		0	-8		-4		0.4	1.00	0.90	1.49	0.15	1.049	0.86	0.14		
		0.0		0	-10		-5		0.5	1.00	0.90	1.65	0.16	1.065	0.85	0.15		
		0.0		0	-12		-6		0.6	1.00	0.90	1.82	0.18	1.082	0.83	0.17		
		0.0		0	-14		-7		0.7	1.00	0.90	2.01	0.20	1.101	0.82	0.18		
		0.0		0	-16		-8		0.8	1.00	0.90	2.23	0.22	1.123	0.80	0.20		
		0.0		0	-18		-9		0.9	1.00	0.90	2.46	0.25	1.146	0.79	0.21		
		0.0		0	-20		-10		1	1.00	0.90	2.72	0.27	1.172	0.77	0.23		

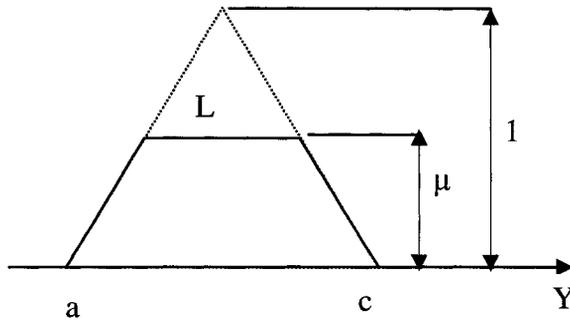
IMPACT OF TRAVEL TIME SAVINGS ON TMS ASSUMING CHANGES TO IVTT AND OVTT																		
Number of buses		IVTT saving is in one direction only - same direction as car travel time is measured																
Orig roundtrip time		60 min																
Original wait time		7.5 min																
Original wait time		7.5 min																
Baseline		Change of utility					BUS OVTT			New Mode Splits								
Mode split		CAR IVTT		BUS IVTT			BUS OVTT		d(u)		Car		Bus		Car+Bus		IVTT +OVTT n=4	
Cars	Bus	d(IVTT)	Coef	d(u)	d(IVTT)	Coef	d(IVTT)	Coef	d(u)	e ^u	P ^{e^u}	e ^u	P ^{e^u}		P ^{car}	P ^{bus}		
0.90	0.10	0.0	-0.025	0	0.0	-0.025	0	-0.05	0	1.00	0.90	1.00	0.10	1.000	0.90	0.10		
		0.0		0	-4		-0.5		0.125	1.00	0.90	1.13	0.11	1.013	0.89	0.11		
		0.0		0	-6		-0.75		0.1875	1.00	0.90	1.21	0.12	1.021	0.88	0.12		
		0.0		0	-8		-1		0.25	1.00	0.90	1.28	0.13	1.028	0.88	0.12		
		0.0		0	-10		-1.25		0.3125	1.00	0.90	1.37	0.14	1.037	0.87	0.13		
		0.0		0	-12		-1.5		0.375	1.00	0.90	1.45	0.15	1.045	0.86	0.14		
		0.0		0	-14		-1.75		0.4375	1.00	0.90	1.55	0.15	1.055	0.85	0.15		
		0.0		0	-16		-2		0.5	1.00	0.90	1.65	0.16	1.065	0.85	0.15		
		0.0		0	-18		-2.25		0.5625	1.00	0.90	1.76	0.18	1.076	0.84	0.16		
		0.0		0	-20		-2.5		0.625	1.00	0.90	1.87	0.19	1.087	0.83	0.17		

Appendix D: Calculation of Transit Modal Share for different TCTTr

Table 13.6: Impact of TCTTr on Transit Modal Share

Existing TMS	Pbus	0.15				
O-D TT	Car TT (min)	60				
O-D TT	Transit TT (min)	120				
Savings	Transit TT saving (m)	5	LL	UL		
IVTT coeff	b		-0.015	-0.026		
OVTT coeff	c		-0.03	-0.05		
existing headway		5				
# of buses	n	24				
	TCTTr	2.5				
	TCTTr ^u	1.917				
	Δu	0.547	0.946			
	P'	0.234	0.313			
			Lower Limit	Upper Limit		
	incr	TCTTr	Δu	P'	Δu	P'
	0.05	2.5	0.000	0.150	0.000	0.150
		2.45	0.047	0.156	0.081	0.161
		2.4	0.094	0.162	0.162	0.172
		2.35	0.141	0.169	0.243	0.184
		2.3	0.187	0.176	0.324	0.196
		2.25	0.234	0.182	0.406	0.209
		2.2	0.281	0.189	0.487	0.223
		2.15	0.328	0.197	0.568	0.237
		2.1	0.375	0.204	0.649	0.252
		2.05	0.422	0.212	0.730	0.268
		2	0.469	0.220	0.811	0.284
		1.95	0.516	0.228	0.892	0.301
		1.9	0.562	0.236	0.973	0.318
		1.85	0.609	0.245	1.055	0.336
		1.8	0.656	0.254	1.136	0.355
		1.75	0.703	0.263	1.217	0.373
		1.7	0.750	0.272	1.298	0.393
		1.65	0.797	0.281	1.379	0.412
		1.6	0.844	0.291	1.460	0.432
		1.55	0.891	0.301	1.541	0.452

Appendix E: Trapezoid Area Calculation



$L=?$

$$\frac{c-a}{2} = \frac{L}{1-\mu}$$

$$(c-a) \cdot (1-\mu) = L$$

Area of the trapezoid is :

$$A = \frac{L + (c-a)}{2} \times \mu = \frac{(c-a) \times (2-\mu)}{2} \times \mu$$

Appendix F: Development of “Partial” Queue Jump

Table 13.7: Signal Timing Parameters at the Bottleneck Intersection

Stage	Movements	Stage duration (sec) (green+clearance)
1	NB LT & SB	30+5=35
2	NB & SB	62+5=67
3	EB	13+5=18
	Total	120

For control development and testing purposes, the traffic volume in the eastbound direction is set so that it continuously oversaturates the bottleneck intersection

Table 13.8: Eastbound Traffic Volumes used for Control Development Purposes

Time	EB Volume (veh/hr)
0-120	1000
120-14400	500

Appendix G: Tests of Hypotheses and Significance

To test whether one control is better than the other the following hypothesis H_0 was used:

$H_0: \mu_1 = \mu_2$, there is essentially no difference between the data

If H_0 can be rejected based on the sample means and standard deviations, one control technique is significantly better than the other. The t statistics for the difference of means is calculated by:

$$t = \frac{|X_1 - X_2|}{\sigma \times \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

where the “pooled” standard deviation is expressed by:

$$\sigma = \sqrt{\frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2 - 2}}$$

X_1, X_2 sample means

s_1, s_2 sample standard deviations

N_1, N_2 sample sizes

σ “pooled” standard deviation

The distribution of t is Student’s distribution with $v = N_1 + N_2 - 2$ degrees of freedom.

The t scores for specific significance levels are shown in the following table. For example, to reject hypothesis H_0 with a 5 % significance level, the calculated t-score has to be more than 1.73. (Source: Spiegel, 1972, pg. 189, 344)

Table 13.9: Student’s t Distribution Percentile Values Used in Significance Test

Degrees of freedom, v	$t_{0.95}$ (Significance level 5%)	$t_{0.99}$ (Significance level 1%)
$10+10-2=18$	1.73	2.55
38	1.68	2.42

Appendix H: “Park View” Type Algorithm Development

Sample Tuning Results

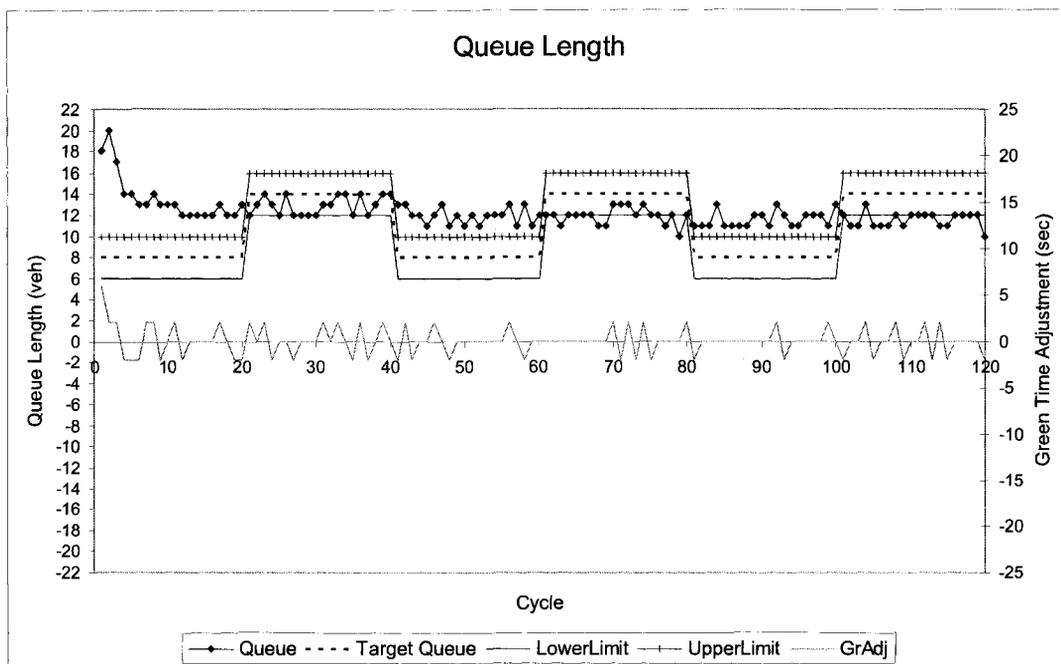


Figure 13.7: “Park View” Algorithm Tuning, Output for Set 1

Table 13.10: “Park View” Algorithm Tuning Parameters and MoP – Set 1

Limits for increment Δ_2^*	GreenAdj. Increments	Accuracy # of points outside $Q_T - \tau$	Responsiveness (Cycles)	RMSE	$\sigma(\text{GrAdj})$
$Q_T \pm 6$	$\Delta_1=2, \Delta_2=4$	64	n/a	3.085	1.245

*Lower limit for increment Δ_2 is $Q_{\text{trig}}=Q_T - 2$ and upper limit for increment Δ_2 is

$Q_{\text{max}}=Q_T - 2$

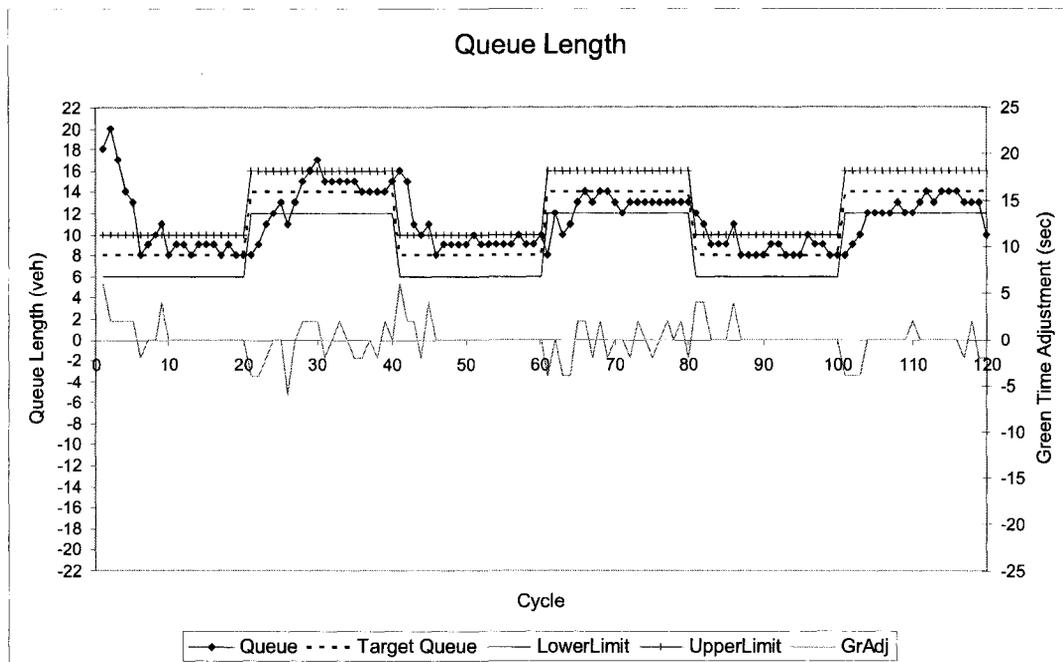


Figure 13.8: “Park View” Algorithm Tuning, Output for Set 2

Table 13.11: “Park View” Algorithm Tuning Parameters and MoP – Set 2

Limits for increment Δ_2	GreenAdj. Increments	Accuracy # of points outside $Q_T \pm \tau$	Responsiveness (Cycles)	RMSE	$\sigma(\text{GrAdj})$
$Q_T \pm 2$	$\Delta_1=2, \Delta_2=4$	21	4,4,2,2,4	2.17	1.846

Appendix I: Fuzzy Logic Tuning for Pre-Signal Operation Control

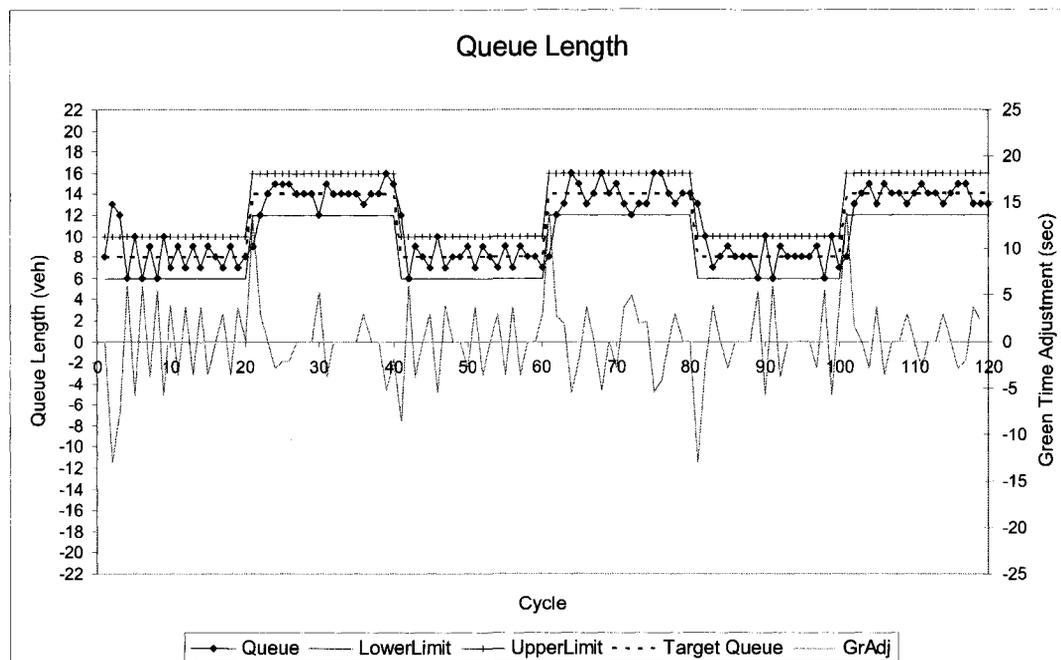


Figure 13.9: Algorithm Tuning, Sample Output for Set 1

Table 13.12: FL Logic Algorithm Tuning Parameters and MoP – Sample Output 1

Max Green Adj. Rmax	Output Scaling factor for output MF,p	Input scaling factors	Accuracy # of points outside $Q_T \pm \tau$	Responsiveness (Cycles)	RMSE	$\sigma(\text{GrAdj})$
16	$p=0.5^*$	sf1=1 sf2=0.4	7	1,1,1,1,1	3.984	1.526

* output MF centres: $c_{-2}=-16$, $c_{-1}=-11$, $c_0=0$, $c_{+1}=11$, $c_{+2}=16$

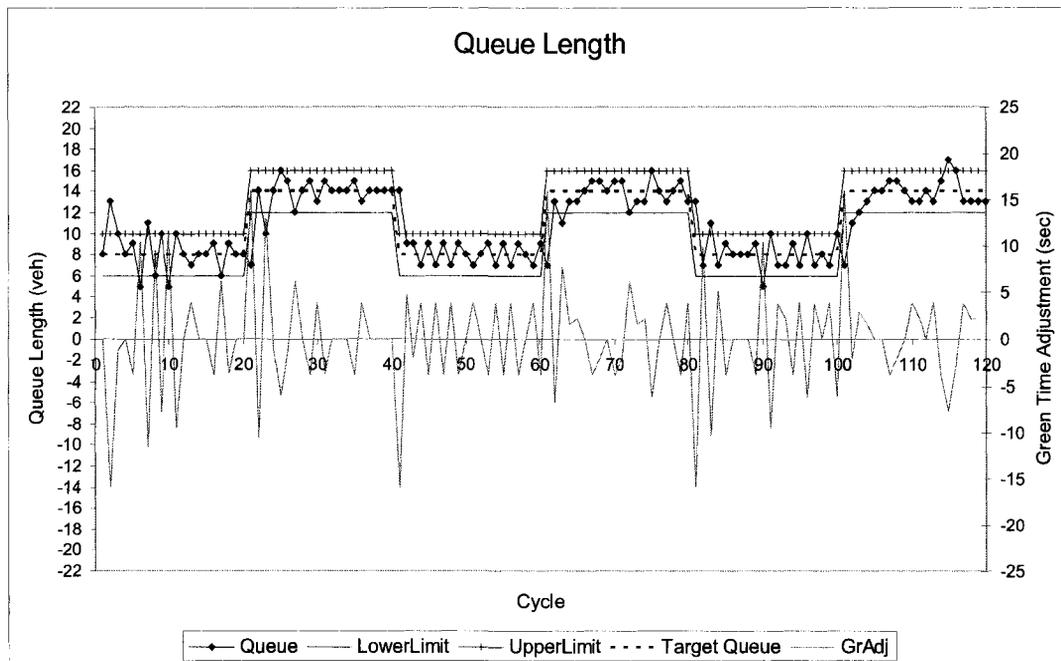


Figure 13.10: FL Algorithm Tuning, Sample Output for Set 2

Table 13.13: FL Logic Algorithm Tuning Parameters and MoP – Sample Output 2

Max Green Adj. Rmax	Output Scaling factor for MF,p	Input scaling factors	Accuracy # of points outside $Q_T \pm \tau$	Responsiveness (Cycles)	RMSE	$\sigma(\text{GrAdj})$
16	$p=0.5^*$	sf1=1 sf2=1.4	16	1,1,1,1,2	3.998	1.634

* output MF centres*: $c_{-2}=-16, c_{-1}=-11, c_0=0, c_{+1}=11, c_{+2}=16$

Appendix J: Simulation Results of Control Algorithm for Pre-Signal Operation

Table 13.14: Travel Times for Individual Simulation Runs for the Base Case Scenario

BASE	Simulation Run	Vehicle TravelTime (sec/veh)		
		Car	Bus	Car+Bus
	1	426.5	438.5	426.7
	2	383.4	395.1	383.6
	3	385.8	390.5	385.9
	4	339.1	352.1	339.3
	5	485.9	486.1	485.9
	6	455.7	467.5	455.9
	7	443.1	456.5	443.3
	8	386.2	403.2	386.5
	9	379	394.5	379.3
	10	435.6	456.5	436.0
	Average	412.0	424.1	412.2
	StDev	44.3	42.8	44.3

Table 13.15: Travel Times for Individual Simulation Runs for the Queue Management without Signal Control Scenario

QMnoSC	Simulation Run	Vehicle TravelTime (sec/veh)		
		Car	Bus	Car+Bus
	1	394.8	209.3	391.5
	2	394.8	200.1	391.3
	3	434.9	200.3	430.7
	4	434.9	200.3	430.7
	5	489.3	196	484.0
	6	510	213.9	504.7
	7	492.7	200.7	487.4
	8	344.4	195.9	341.7
	9	387.2	204.8	383.9
	10	508.9	213.8	503.6
	Average	439.2	203.5	434.9
	StDev	58.6	6.7	57.6

Table 13.16: Travel Times for Individual Simulation Runs for the “Park View” Control Scenario

ParkView	Simulation Run	Vehicle TravelTime (sec/veh)		
		Car	Bus	Car+Bus
	1	442.3	155.2	437.1
	2	381.4	160.6	377.4
	3	403	142.2	398.3
	4	536.7	156.1	529.8
	5	548.5	147	541.3
	6	406.2	151.7	401.6
	7	441.3	155.5	436.1
	8	370.4	151.5	366.5
	9	503.3	150.7	496.9
	10	310.3	151.1	307.4
	Average	434.3	152.2	429.3
	StDev	76.3	5.1	74.9

Table 13.17: Travel Times for Individual Simulation Runs for the “On-Off” Control Scenario

ON-OFF	Simulation Run	Vehicle TravelTime (sec/veh)		
		Car	Bus	Car+Bus
	1	451.7	140.5	446.1
	2	350.5	135.2	346.6
	3	376.2	140.1	371.9
	4	459.2	140.3	453.5
	5	510.3	140.3	503.6
	6	498.7	140.2	492.2
	7	415.4	140.3	410.4
	8	399.6	136	394.8
	9	350.6	139.9	346.8
	10	415.8	140.3	410.8
	Average	422.8	139.3	417.7
	StDev	56.6	2.0	55.6

Table 13.18: Travel Times for Individual Simulation Runs for the Model Based Control Scenario

MODEL	Simulation Run	Vehicle TravelTime (sec/veh)		
		Car	Bus	Car+Bus
	1	410.1	141.7	405.3
	2	393.5	141.2	389.0
	3	377.3	141.8	373.1
	4	408.2	141.9	403.4
	5	506.3	141.7	499.7
	6	444.8	141.3	439.3
	7	426	141	420.9
	8	448.6	140.7	443.1
	9	364.4	141.4	360.4
	10	441.4	141.8	436.0
	Average	422.1	141.5	417.0
	StDev	41.1	0.4	40.3

Table 13.19: Travel Times for Individual Simulation Runs for the Fuzzy Logic Control Scenario

FL	Simulation Run	Vehicle TravelTime (sec/veh)		
		Car	Bus	Car+Bus
	1	417.2	144.7	412.3
	2	421.4	136.7	416.3
	3	416.8	140.9	411.8
	4	443	136.6	437.5
	5	533.3	141.2	526.2
	6	402.2	145.3	397.6
	7	376.5	140.6	372.2
	8	357.9	140.9	354.0
	9	463.9	136.8	458.0
	10	362	140.4	358.0
	Average	419.4	140.4	414.4
	StDev	52.4	3.1	51.5

Table 13.20: Statistical Comparison of the Base Case and QMnoSC Scenarios

	BASE		QMnoSC		Difference	t-score	Significance level	
	Aver	STDev	Aver	STDev			0.05	0.01
Car	412.0	44.3	439.2	58.6	27.2	1.11	No	No
Bus	424.1	42.8	203.5	6.7	-220.5	-15.28	Yes	Yes
Car+Bus	412.2	44.3	434.9	57.6	22.7	0.95	No	No

Table 13.21: Statistical Comparison of the Base Case and “Park Vies” Control Scenarios

	BASE		ParkView		Difference	t-score	Significance level	
	Aver	STDev	Aver	STDev			0.05	0.01
Car	412.0	44.3	434.3	76.3	22.3	0.76	No	No
Bus	424.1	42.8	152.2	5.1	-271.9	-18.93	Yes	Yes
Car+Bus	412.2	44.3	429.3	74.9	17.0	0.59	No	No

Table 13.22: Statistical Comparison of the Base Case and “On-Off” Control QMnoSC Scenarios

	BASE		ON-OFF		Difference	t-score	Significance level	
	Aver	STDev	Aver	STDev			0.05	0.01
Car	412.0	44.3	422.8	56.6	10.8	0.45	No	No
Bus	424.1	42.8	139.3	2.0	-284.7	-19.94	Yes	Yes
Car+Bus	412.2	44.3	417.7	55.6	5.4	0.24	No	No

Table 13.23: Statistical Comparison of the Base Case and Model Based Control Scenarios

	BASE		MODEL		Difference	t-score	Significance level	
	Aver	STDev	Aver	STDev			0.05	0.01
Car	412.0	44.3	422.1	41.1	10.0	0.50	No	No
Bus	424.1	42.8	141.5	0.4	-282.6	-19.8	Yes	Yes
Car+Bus	412.2	44.3	417.0	40.3	4.8	0.29	No	No

Table 13.24: Statistical Comparison of the Base Case and Fuzzy Logic Based Control Scenarios

	BASE		FL		Difference	t-score	Significance level	
	Aver	STDev	Aver	STDev			0.05	0.01
Car	412.0	44.3	419.4	52.4	7.4	0.32	No	No
Bus	424.1	42.8	140.4	3.1	-283.6	-19.83	Yes	Yes
Car+Bus	412.2	44.3	414.4	51.5	2.1	0.10	No	No

Appendix K: Background Data for the Carling Avenue Example

Carling and Bronson

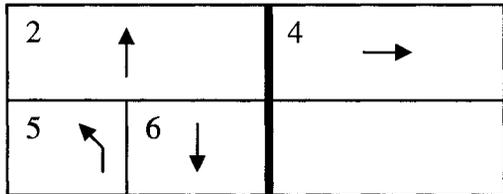


Figure 13.11: Timing Sequence at Carling and Bronson

Table 13.25: Signal Group Sequencing at Carling and Bronson

Signal Group	Timing
2	103
4	27
5	26
6	77
2-6 P**	7+V+14
4-8 P	7+14
Offset	46
Cycle	130

Table 13.26: Approximate Capacity Calculation

Carling Bronson

Current configuration

Cycle	130	NB ST 2	EB ST 4	NB LT 5	SB ST 6	EB LT 7	Cycle Length Check	
							Ring A	Ring B
SF		1800	1500	1800	1800	1000		
VOL		1130	99	263	1429	235		
Phase		103	27	26	77	27	130	130
LostTime		6	6	6	6	6		
Lanes		1.5	0.5	1	2	1.5		
Green		97	21	20	71	21		
Capacity		2015	121	277	1966	242		
Manual V/C		0.56	0.82	0.95	0.73	0.97		

Carling and Preston

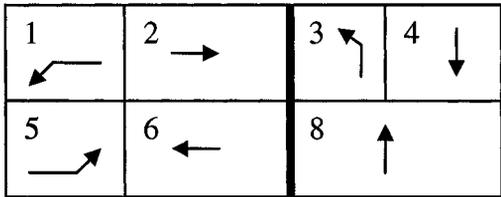


Figure 13.12: Timing Sequence at Carling and Preston

Table 13.27: Signal Group Sequencing at Carling and Preston

Signal Group	Timing
1	24
2	27
3	30
4	39
5	24
6	27
8	69
2-6 P	7+V+14
4-8 P	19+9
Offset	6
Cycle	120

Table 13.28: Approximate Capacity Calculation

Carling Preston

Current configuration

Cycle	120	WB LT	EBST	NB LT	SB ST	EB LT	WB ST	NB ST	Cycle Length	
									Check	Ring B
Signal group		1	2	3	4	5	6	8	Ring A	Ring B
SF		1900	1900	1900	1900	1900	1900	1900		
VOL		287	463	278	487	148	894	1022		
Phase		24	27	30	39	24	27	69	120	120
LostTime		8	6	8	8	6	6	6		
Lanes		1	3	1	2	2	3	2		
Green		18	21	24	33	18	21	63		
Capacity		285	998	380	1045	570	998	1995		
Manual V/C		1.01	0.46	0.73	0.47	0.26	0.90	0.51		

Carling and Booth

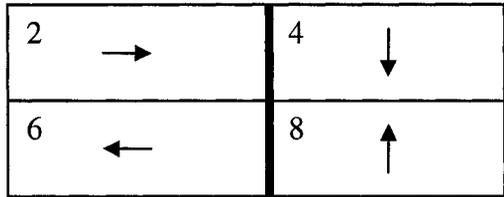


Figure 13.13: Timing Sequence at Carling and Booth

Table 13.29: Signal Group Sequencing at Carling and Booth

Signal Group	Timing
2	74
4	46
6	74
8	76
2-6 P**	15+V+8
4-8 P	17+8
Offset	101
Cycle	120

V/C calculation was not carried out because this is obviously not a critical intersection

Appendix L: Carling Avenue Evaluation

Table 13.30: Vehicle Travel Perspective: Vehicle Travel Time Difference

Statistical Test: Vehicle Travel Time Change

		n		10	Student's t distribution D.F.=10+10-2=18			
		Before	Scen A	Diff	Pooled StDev	t	1.73 t(.95)	2.55 t(.99)
Car	TT	420	443	22				
	StDev	54	39		49.6	1.01	Not Sign	Not Sign
Bus	TT	401	263	-139				
	StDev	87	5		65.3	4.75	Sign	Sign
Car+Bus	TT	420	440	20				
	StDev	53	38		49.1	0.91	Not Sign	Not Sign

Statistical Test: Vehicle Travel Time Change

		n		10	Student's t distribution D.F.=10+10-2=18			
		Scen A	Scen B	Diff	Pooled StDev	t	1.73 t(.95)	2.55 t(.99)
Car	TT	443	432	-10				
	StDev	39	39		41.3	0.57	Not Sign	Not Sign
Bus	TT	263	213	-50				
	StDev	5	5		5.3	21.14	Sign	Sign
Car+Bus	TT	440	429	-11				
	StDev	38	39		40.6	0.61	Not Sign	Not Sign

Statistical Test: Vehicle Travel Time Change

		n		10	Student's t distribution D.F.=10+10-2=18			
		Before	Scen B	Diff	Pooled StDev	t	1.73 t(.95)	2.55 t(.99)
Car	TT	420	432	12				
	StDev	54	39		49.7	0.54	Not Sign	Not Sign
Bus	TT	401	213	-188				
	StDev	87	5		65.3	6.45	Sign	Sign
Car+Bus	TT	420	429	9				
	StDev	53	39		49.2	0.40	Not Sign	Not Sign

Table 13.31: Person Travel (Bimodal) Perspective: Person Travel Time Difference

Statistical Test: Pedestrian Travel Time Change								
		n			Student's t distribution D.F.=10+10-2=18			
		Before	Scen A	Diff	Pooled StDev	t	t(.95)	t(.99)
CarPas	TT	420	443	22				
	StDev	54	39		49.6	1.01	Not Sign	Not Sign
BusPas	TT	401	263	-139				
	StDev	87	5		65.3	4.75	Sign	Sign
Car+Bus	TT	414	380	-34				
Pas	StDev	53	25		43.8	1.74	Sign	Not Sign

Statistical Test: Pedestrian Travel Time Change

Statistical Test: Pedestrian Travel Time Change								
		n			Student's t distribution D.F.=10+10-2=18			
		Scen A	Scen B	Diff	Pooled StDev	t	t(.95)	t(.99)
CarPas	TT	443	432	-10				
	StDev	39	39		41.3	0.57	Not Sign	Not Sign
BusPas	TT	263	213	-50				
	StDev	5	5		5.3	21.14	Sign	Sign
Car+Bus	TT	380	355	-25				
Pas	StDev	25	26		27.0	2.06	Sign	Not Sign

Statistical Test: Pedestrian Travel Time Change

Statistical Test: Pedestrian Travel Time Change								
		n			Student's t distribution D.F.=10+10-2=18			
		Before	Scen B	Diff	Pooled StDev	t	t(.95)	t(.99)
CarPas	TT	420	432	12				
	StDev	54	39		49.7	0.54	Not Sign	Not Sign
BusPas	TT	401	213	-188				
	StDev	87	5		65.3	6.45	Sign	Sign
Car+Bus	TT	414	355	-59				
Pas	StDev	53	26		43.9	3.00	Sign	Sign

Appendix M: Transit Volumes in the Woodroffe Avenue Corridor

Twelve transit routes operate in the analysed area. Ten routes travel in the North and South directions at least partially along Woodroffe Avenue. The other two routes travel East-West and were included to properly simulate signal operation. The 12 transit routes were grouped into three categories (Line A, Line B, and Line C) based on their travel paths (Table 13.33). The paths are depicted in Figure 13.14.

The numbers represent the mean hourly number of transit vehicles during the morning peak period between 7 AM and 9 AM.

Table 13.32: Transit Route Frequencies

Route # on Line A	Number of buses (veh/hr)	
	NB	SB
70	4	0
71	4	0
73	4	0
76	5	0
77	6	0
95	18	18
Total:	41	18
Route # on Line B	Number of buses (veh/hr)	
	NB/WB	SB/EB
57	0	2
86	3	8
111	4	4
116	3	3
Total:	10	17
Route # on Line C	Number of buses (veh/hr)	
	EB	WB
55	2	0
118	8	8
Total:	10	8

Table 13.33: Transit Lines and Volumes (Hourly Number of Transit Vehicles During the two –hour Morning Peak Period)

Segment	From	To	Frequency (bus/hr)
Line A NB	A	F	41
Line A SB	F	A	18
Line B NB/WB	C	F	10
Line B SB/EB	F	C	17
Line C EB	G	F	8
Line C WB	F	G	18

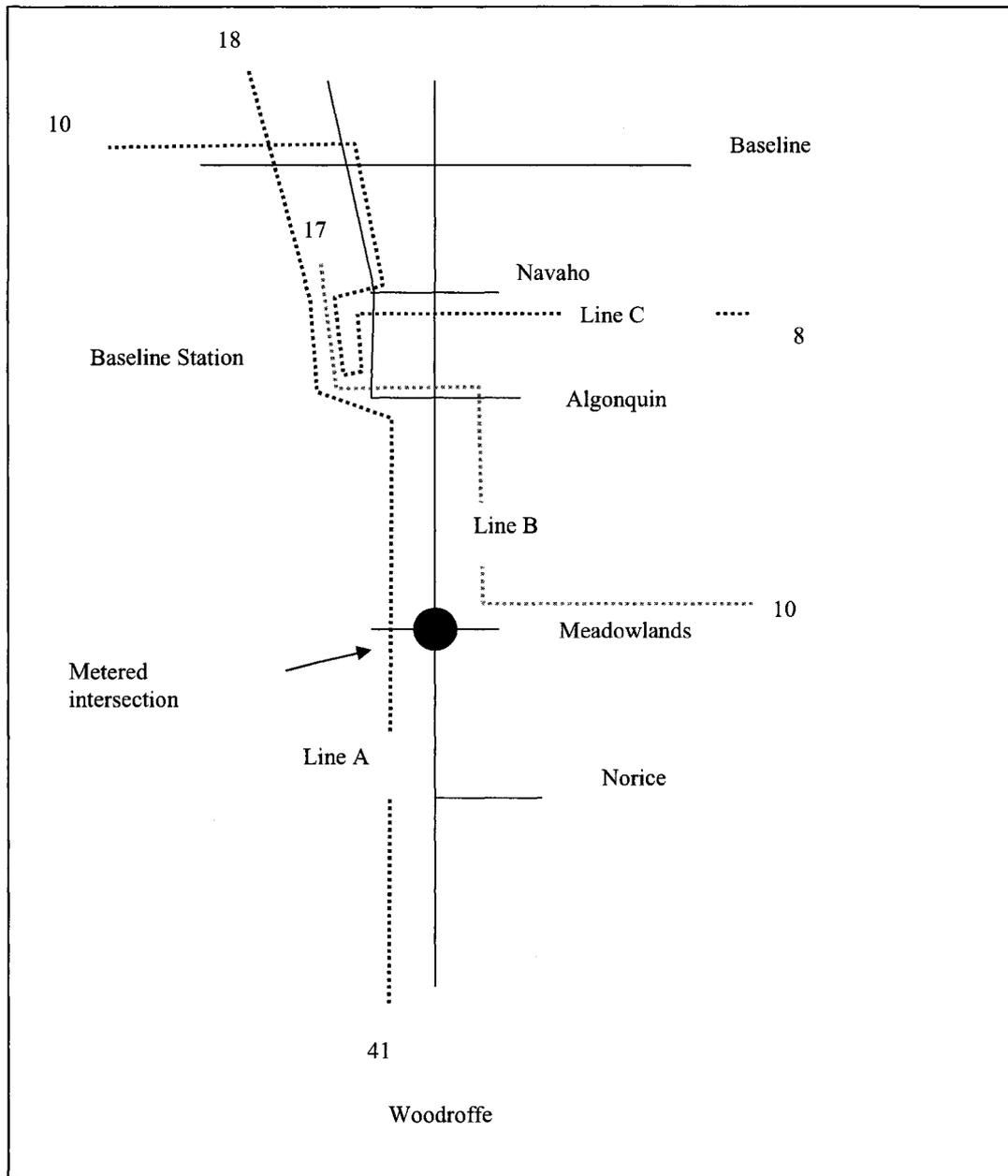


Figure 13.14: Transit Frequencies by Transit Lines

Appendix N: Volume Balancing

Volume balancing is required when the conservation of traffic flows principle is not satisfied. This often happens when traffic counts at individual intersections were not taken during the same time and day. When volumes counts are taken on different days, the number of vehicles leaving an upstream intersection is likely to be different from the number of vehicles approaching the adjacent downstream intersection.

There is no unique solution to the balancing problem thus there is no standard procedure for carrying out volume balancing. The procedure presented in this section was developed for this research and it is reasonable accurate for the purpose of this research. The first part of volume balancing is through Step 1 to Step 8 (Figure 13.15) where the relationships between the entrance and exit volumes are established for each intersection, and volumes departing upstream intersection are equalized with the approach volumes to downstream intersections. This ensures that vehicles do not appear or disappear between intersections (Conservation of flow principle).

The first part of the balancing process anchors all volumes to the entry volume at the most upstream intersection. Thus, only at the most upstream intersection will approach and turning volumes is identical to the volume counts determined in Step 1. For all other downstream intersections, approach volumes will likely differ from the volumes collected in Step 1. For example at Norice, the most upstream external link, the difference is 0 but at Algonquin, the difference is 310 vehicles (Table 13.34).

The purpose of the last step, Step 9, is to select an “appropriate” entry volume at the most upstream external link. At least two alternatives are available for this selection:

- the entry volume at the up most external link is selected to minimize the total difference between the approach volumes at all intersections; and
- an intersection within the simulated corridor is selected as an anchor point whose traffic counts are assumed to be representative (correct), and the upstream and downstream volumes are calculated based on this anchor point

In the absence of a specific intersection for which it could have been assumed that the collected traffic volumes were representative, the approach of calculating the minimum total difference was adopted. Table 13.34 and Table 13.35 summarize the balancing process. Figure 9.13 depicts the original traffic counts and the calculated percentages of turning movements. Figure 13.17 shows the calculated volumes through Step 1 to Step 8, and the final balanced volume calculated through Step 9.

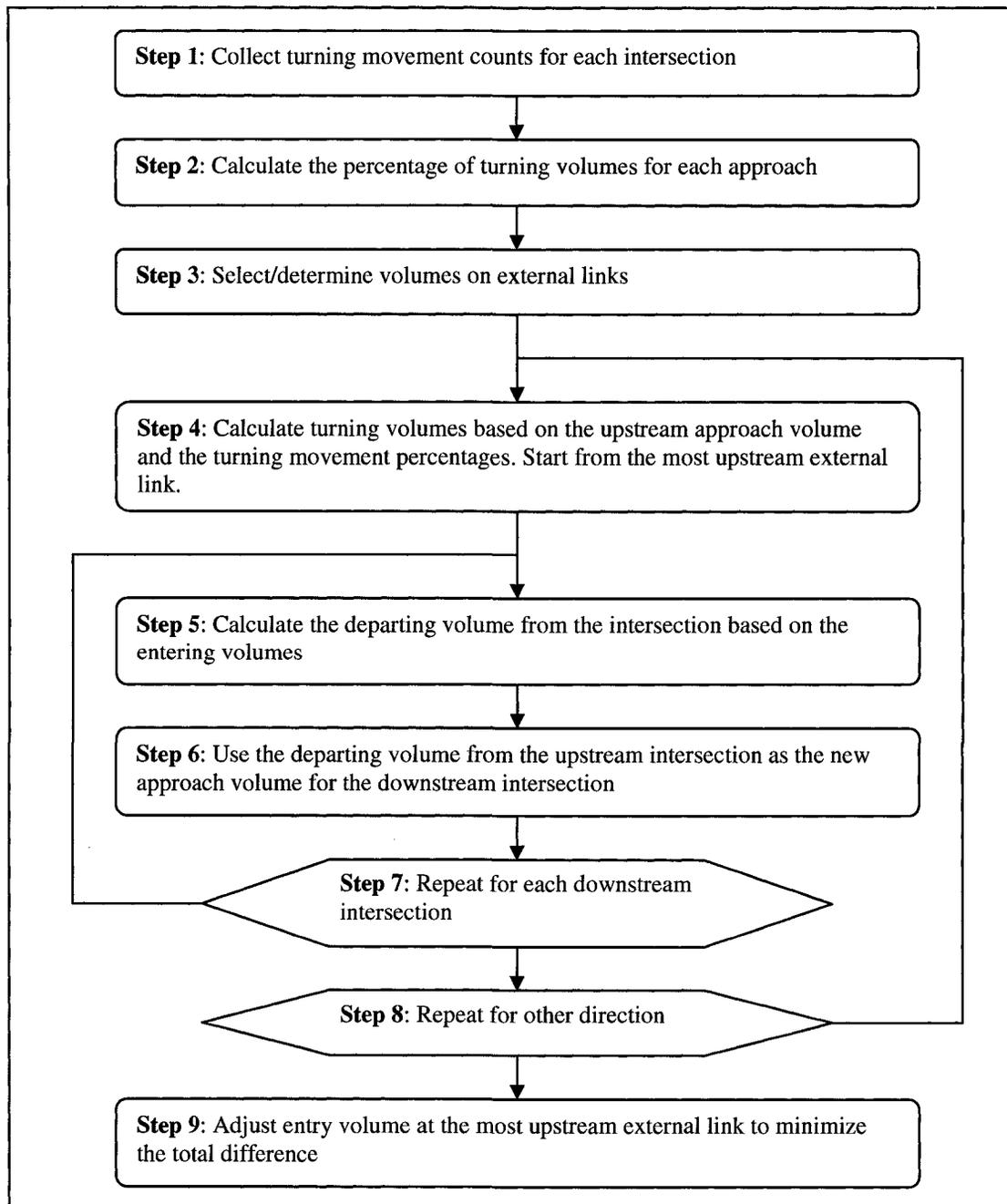


Figure 13.15: The Volume Balancing Process

Table 13.34: Volume Balancing Summary for the NB Direction

NB	Volume Count	Step 1 to Step 9 Volumes	Diff. 1	Step 9 Balanced volumes	Diff. 2
Baseline	1334	1143	191	1256	78
Navaho	1630	1370	260	1512	118
Algonquin	1871	1561	310	1730	141
Meadowlands	2207	2043	164	2293	-86
Norice	2024	2024	0	2274	-250
Total:			925		1

Table 13.35: Volume Balancing Summary for the SB Direction

SB	Volume Count	Step 1 to Step 9 Volumes	Diff. 1	Step 9 Balanced volumes	Diff. 2
Baseline	1145	1145	0	1353	-208
Navaho	837	651	186	738	99
Algonquin	837	705	132	778	59
Meadowlands	804	687	117	750	54
Norice	782	739	43	786	-4
Total:			478		0

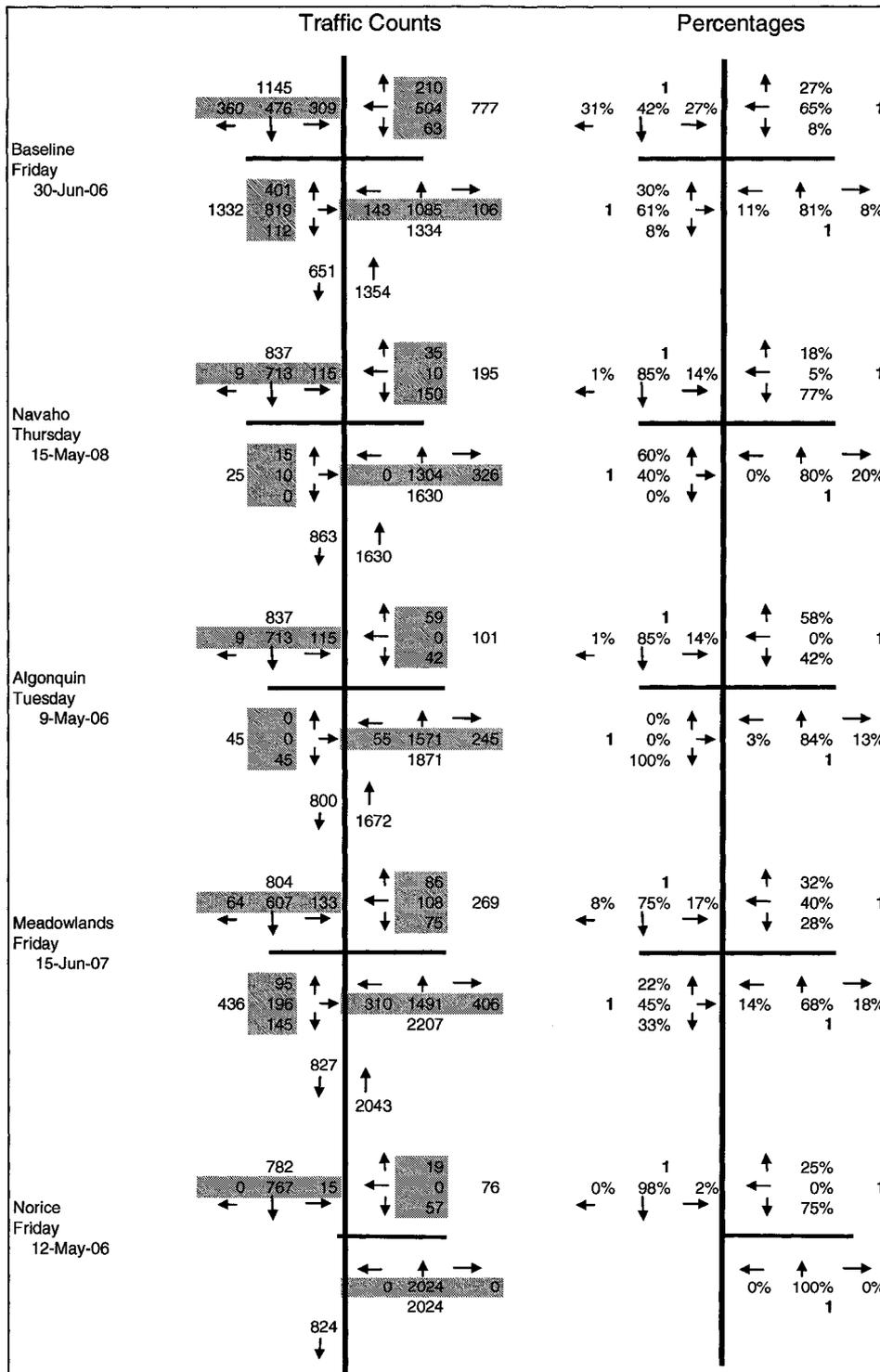


Figure 13.16: Traffic Counts and Turning Movement Percentages

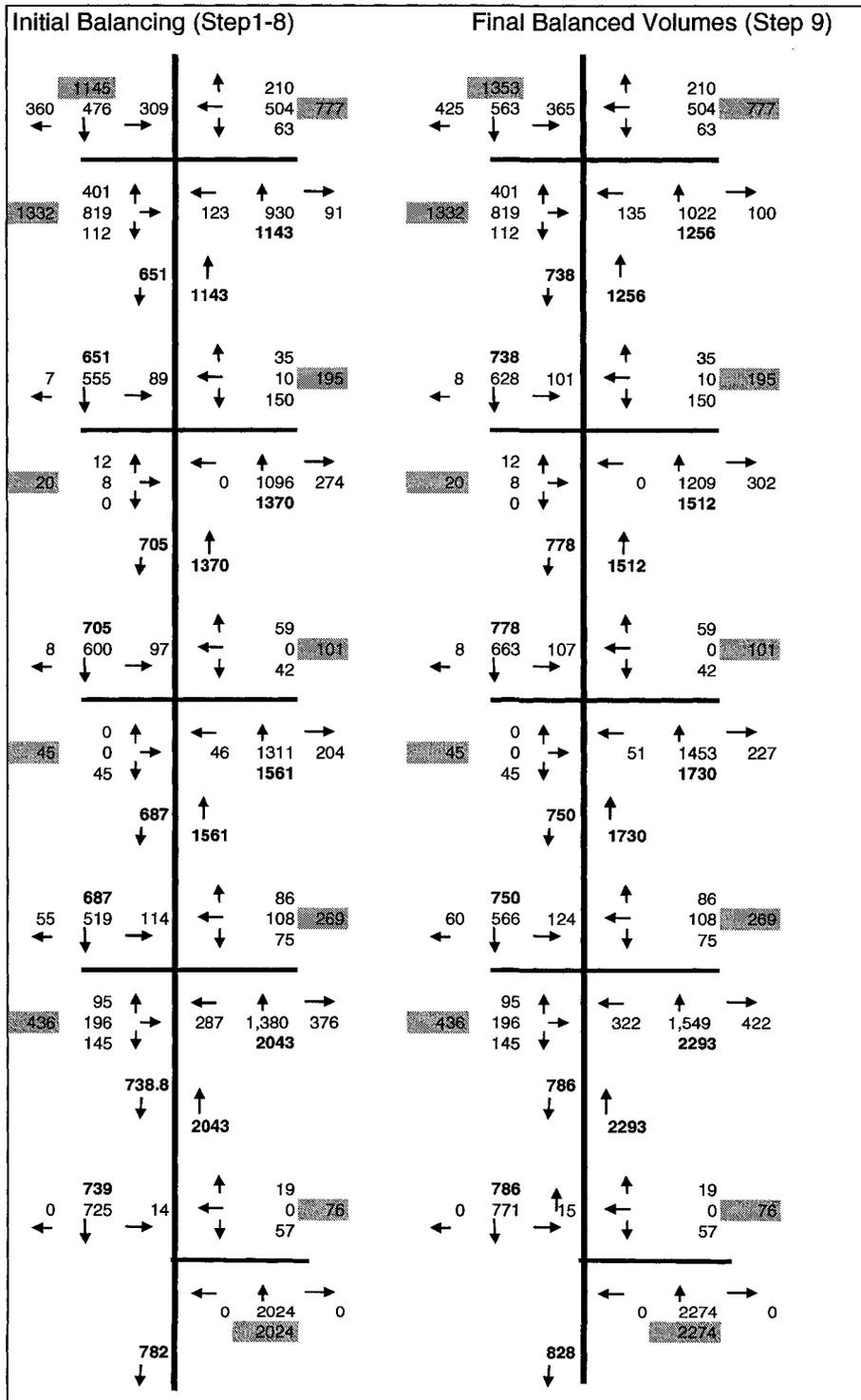


Figure 13.17: Balanced Volumes Through Steps 1 to 8 and Step 9

Appendix O: Optimization of Signal Timing in the Woodroffe Corridor

The procedure used for signal timing optimization is presented in Figure 13.18:

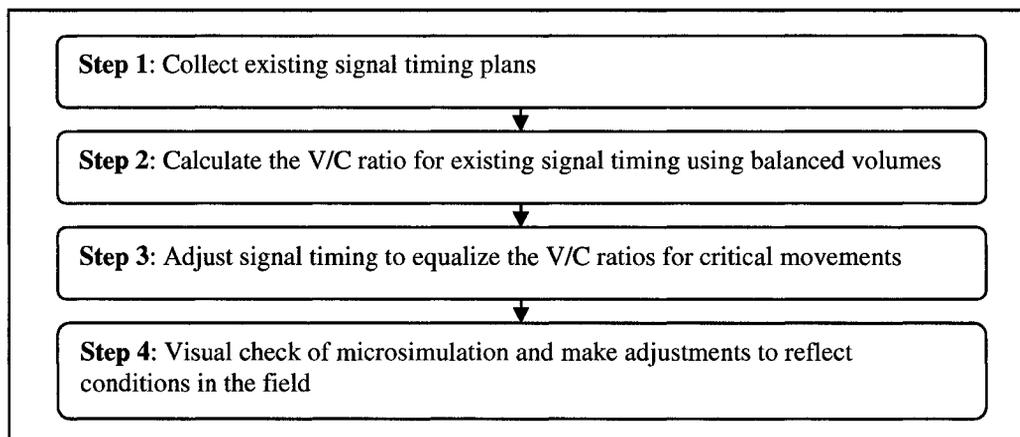


Figure 13.18: Manual Optimization of Signal Splits

Step 1, Step 2, and Step 3:

Signal timing parameters and the results of the optimization steps are summarized individually for each intersection. For example, for the intersection of Woodroffe Avenue and Baseline, the signal group sequencing is provided in Figure 13.19. Table 13.36 summarizes signal timing parameters provided by the City (2nd column), timing parameters calculated by the manual optimization (Step 3) and signal timing parameters implemented in the simulation. Table 13.37 shows the input and output values used in the V/C calculation. (omitted for obviously non-critical intersections).

Step 4 is equivalent to the “fine tuning” carried out by traffic engineers by observing traffic patterns and signal timing in the field. This time, simulation visualization substitutes the field observation. The final signal timing parameters are presented in the 4th column of Table 13.36.

Appendix P: Signal Timing Parameters and the V/C Calculation

Woodroffe and Baseline

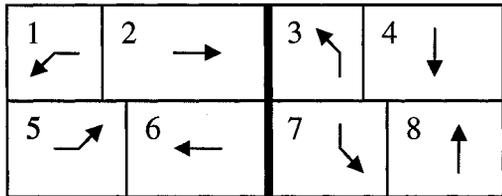


Figure 13.19: Timing Sequence at Woodroffe and Baseline

Table 13.36: Signal Group Sequencing at Woodroffe and Baseline

Signal Group	Timing		
	Provided by City	Calculated for Balanced Volumes	Adjusted through Simulation
1*	15	12	12
2	57	52	52
3	12	16	16
4	46	50	50
5	36	32	32
6	36	32	32
7	18	24	30
8	40	42	36
2-6 P**	7+V+17	No change	
4-8 P	7+29	No Change	
Offset	16	No Change	

* For vehicle signal groups the duration represents green time plus clearance interval

** For pedestrian signal groups, the duration includes minimum green time, variable interval (V) for the coordinated signal groups, and the clearance interval (Flashing don't walk)

Table 13.37: V/C calculation for Balanced Volumes

Woodroffe Baseline

Current configuration		WB LT				EB ST				NB LT				SB ST				Cycle Length	
Cycle	130	1	2	3	4	5	6	7	8	5	6	7	8	5	6	7	8	Check	
Signal group		1	2	3	4	5	6	7	8	5	6	7	8	5	6	7	8	Ring A	Ring B
SF		1900	1900	1900	1900	1900	1900	1900	1900	1900	1900	1900	1900	1900	1900	1900	1900		
VOL		63	819	135	563	401	504	365	1022	401	504	365	1022	401	504	365	1022		
Phase		12	52	16	50	32	62	24	42	32	62	24	42	32	62	24	42	130	130
LostTime		6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6		
Lanes		1	2	2	2	2	3	2	2.2	2	3	2	2.2	2	3	2	2.2		
Green		6	46	10	44	26	26	18	36	26	26	18	36	26	26	18	36		
Capacity		88	1345	292	1286	760	1140	526	1158	760	1140	526	1158	760	1140	526	1158		
Manual V/C		0.72	0.61	0.46	0.44	0.53	0.44	0.69	0.88	0.53	0.44	0.69	0.88	0.53	0.44	0.69	0.88		

Note: There are 3 NB ST lanes but the curb lane ends at the far side of the intersection and it is not fully utilized

Woodroffe and Navaho

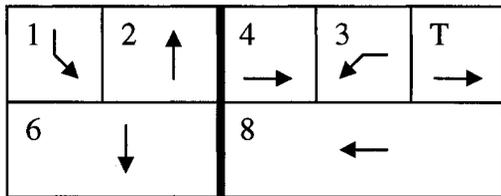


Figure 13.20: Timing Sequence at Woodroffe and Navaho

Table 13.38: Signal Timing at Woodroffe and Navaho

Signal Group	Timing	
	Provided by City	Adjusted through Simulation
1*	15	15
2	65	65
4	32	32
3	10	10
T	8	8
6	80	80
8	50	50
2-6 P	7+V+21	No change
4-8 P	7+28	No change
Offset		No Change

* protected plus permissive left turn

V/C calculation for Woodroffe and Navaho was not carried out because this is obviously not a critical intersection

Woodroffe and Algonquin

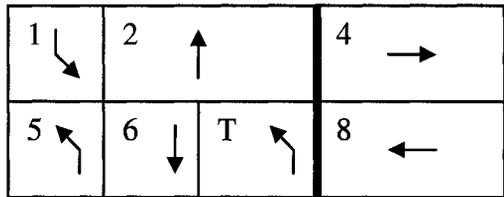


Figure 13.21: Timing Sequence at Woodroffe and Algonquin

Table 13.39: Signal Timing at Woodroffe and Algonquin

Signal Group	Timing		
	Provided by City	Calculated for Balanced Volumes	Adjusted through Simulation
1	14	14	14
2	77	77	77
4	39	39	39
5	14	30*	14
6	61	61	61
T	16	n/a	16
8	39	39	39
2-6 P	10+V+16	No change	No change
4-8 P	12+27	No Change	No change
Offset	27	No Change	No Change

*Includes the transit phase with two clearance intervals

Table 13.40: V/C Calculation for Balanced Volumes

Woodroffe Algonquin

Current configuration		SB LT	NB ST	EB LT	NB LT	SB ST	WB LT	Cycle Length	
Cycle	130	1	2	4	5	6	8	Check	
								Ring A	Ring B
SF		1900	1900	1900	1900	1900	1900		
VOL		107	1453	42	55	663	42		
Phase		14	77	39	30	61	39	130	130
LostTime		6	6	6	12	6	6		
Lanes		2	2	1	1	2	1		
Green		8	71	33	18	55	33		
Capacity		234	2075	482	263	1608	482		
Manual V/C		0.46	0.70	0.09	0.21	0.41	0.09		

The 'lost time' or clearance interval for the NB LT signal group is 12 sec (6x2) because it is displayed twice in the cycle

Woodroffe and Meadowlands

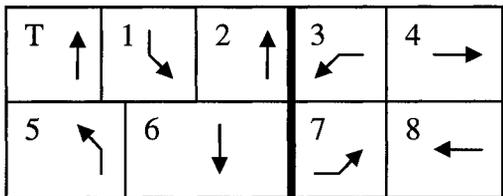


Figure 13.22: Implemented Timing Sequencing at Woodroffe and Meadowlands

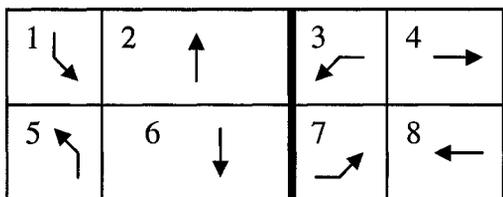


Figure 13.23: Simulated Timing Sequencing at Woodroffe and Meadowlands

Table 13.41: Signal Timing at Woodroffe and Meadowlands

Signal Group	Timing		
	Provided by City	Calculated for Balanced Volumes	Adjusted through Simulation
T	6	n/a	n/a
1	21	16	18*
2	52	66	64
3	15	12	12
4	36	36	36
5	17	24	18
6	62	58	64
7	15	12	12
8	36	36	36
2-6 P	7+V+17	No change	
4-8 P	7+29	No Change	
Offset	2	No Change	

Time of sg_T is added to the time of sg_1

Table 13.42: V/C Calculation for Balanced Volumes

Woodroffe Meadowlands

Current configuration Cycle	SB LT				NB ST				WB ST				Cycle Length Check Ring A	Ring B
	1	2	3	4	5	6	7	8						
SF	1900	1900	1900	1900	1900	1900	1900	1900	1900					
VOL	124	1549	75	196	322	566	75	108						
Phase	16	66	12	36	24	58	12	36						
LostTime	6	6	6	6	6	6	6	6						
Lanes	1	2	2	1	2	2	1	1						
Green	10	60	6	30	18	52	6	30						
Capacity	146	1754	175	438	526	1520	88	438						
Manual V/C	0.85	0.88	0.43	0.45	0.61	0.37	0.86	0.25						

Woodroffe and Norice

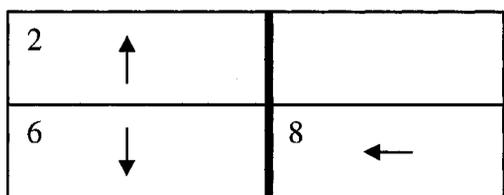


Figure 13.24: Timing Sequence at Woodroffe and Norice

Table 13.43: Signal Timing at Woodroffe and Norice

Signal Group	Timing	
	Provided by City	Adjusted through Simulation
2	97	97
6	97	97
8	33	33
2-6 P	8+V+17	No Change
4-8 P	7+26	No Change
Offset	95	No Change

V/C calculation not carried out because this is not a critical intersection

Appendix Q: Saturation Flow Calibration in VISSIM

There are a number of slightly different methods for collecting and measuring saturation flow data. The HCM 2000 method starts counting from the fourth vehicle that crosses the stop bar. The Australian method (Akcelik, 1993) starts counting vehicles after ten seconds. The reason for not counting vehicles from the very beginning of the green interval is to avoid the impact of start-up delay. Both methods stop counting vehicles when the initial queue has dissipated. Then the mean headway between vehicles and the corresponding hourly volume is calculated. The Canadian Capacity Guide (CCG, 1995) method calculates the headways starting from the first vehicle until the queue dissipates. This method captures the start-up delay thus it would show somewhat lower saturation flow rates compared to the HCM and the Australian methods.

For VISSIM calibration, the CCG was used because the outcomes of the data collections, both from the field study and VISSIM are more relevant since they include the impact of start-up delay. The process of data collections in the field and from VISSIM was identical. The number of vehicles that passed the stop bar during the entire saturated green interval was collected, and the saturation flow rate was calculated by using 13.1 and 13.2.

$$s(i) = \frac{V_{cyc}(i)}{g(i)} \times 3600 \quad 13.1$$

$$s = \frac{1}{n} \sum_n s(i) \quad 13.2$$

In VISSIM, the saturation flow cannot be explicitly defined (VISSIM, 2007). Saturation flow is modified by changing “driver parameters” in the Wiedemann 74 car following model. In the model, the distance between two vehicles is calculated by:

$$d = ax + bx \quad 13.3$$

$$bx = (bx_add + bx_mult \times z) \times \sqrt{v} \quad 13.4$$

d – distance between two vehicles (m)

ax – mean standstill distance (m)– defines the mean desired distance between stopped cars

bx_add – additive part of the desired safety distance

bx_mult – multiplicative part of the safety distance

v – vehicle speed (m/s)

z – normally distributed random number in the rang of 0 to 1 with a mean of 0.5 and standard deviation of 0.15

Based on Equation 13.4, the distance between two vehicles has two components. The first component, ax , specifies the minimum distance between two vehicles that are stopped.

The second component describes the distance between two vehicles when they are moving and this component increases with the square root of the speed. The second component has two parameters, bx_add , and bx_mult . These two parameters are normally modified to change the saturation flow. As these two parameters increase, the distance

between two vehicles increases, thus the headway and saturation flow rate decrease.

Instead of changing these two parameters independently, the VISSIM Manual (VISSIM, 2006) suggests using the following relationship:

$$bx_mult = bx_add + 1$$

13.5

Figure 13.25 summarizes saturation flow rates measured in VISSIM for the intersection of Woodroffe and Meadowlands with various values of bx_mult .

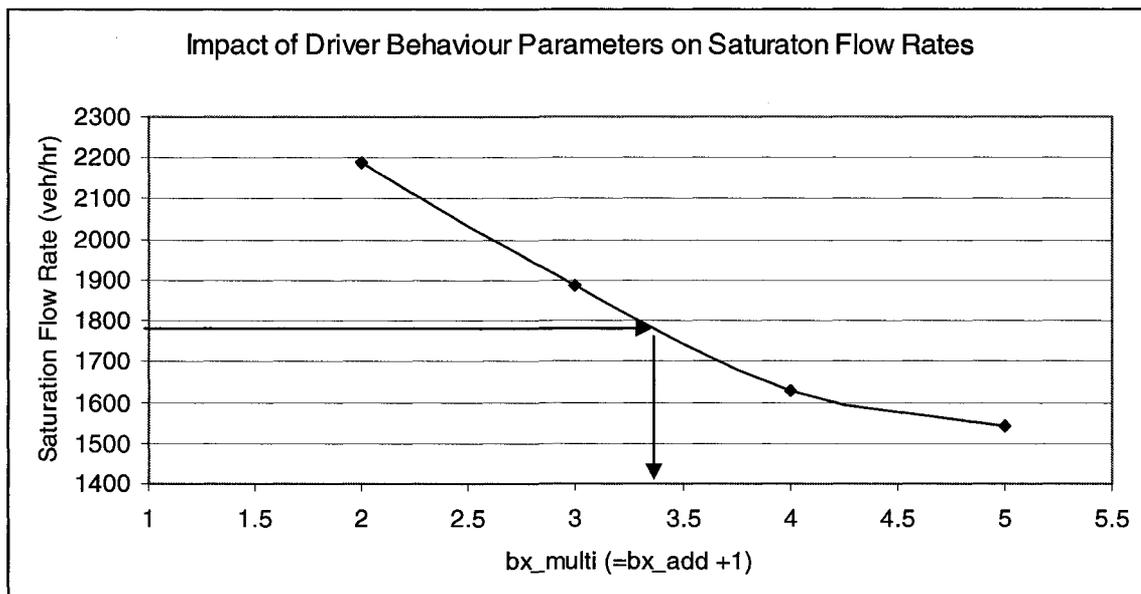


Figure 13.25: Impact of driver behaviour on the saturation flow rate

The saturation flow rate measured at the intersection of Woodroffe and Meadowlands for the straight through movement in the northbound direction was 1780 veh/hr (Table 13.44). The corresponding driver behaviour parameters for the VISSIM simulation are $bx_mult=3.4$ and $bx_add=2.4$. It was assumed that these values are applicable for all other locations and all other movements on the simulated road sections.

Table 13.44: Saturation Flow Rate Measured in the Field

Location	Date and Time	Weather Conditions	Saturation Flow Rate (Veh/hr)
Woodroffe NB ST at Meadowlands	May 27, 2008 8:10-8:40 AM	Sunny, dry pavement	1780

Appendix R: Traffic Count Profiles in the Woodroffe Corridor

Table 13.45: Distribution of Arrival Volumes at Various Locations

	Knoxdale Depart	Norice ST	Meadowlands ST	Algonquin ST
7:00	265	283	309	313
7:15	317	411	369	345
7:30	398	490	360	415
7:45	404	527	394	323
8:00	369	492	375	428
8:15	400	530	357	405
8:30	200	524	365	416
8:45	352	441	287	350
9:00	206	382	261	320
9:15	259	332	273	284
9:30	222	243	204	275

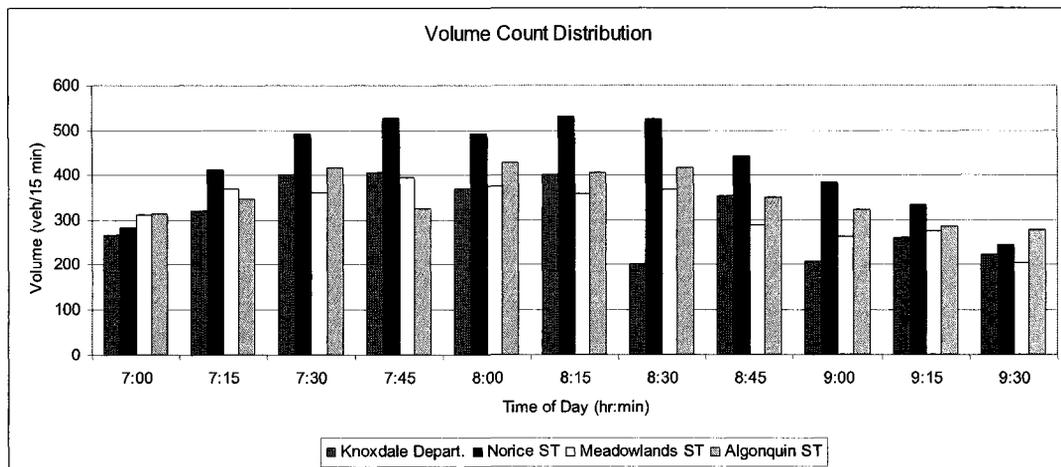


Figure 13.26: Distribution of Arrival Volumes at Various Locations

Appendix S: Simulation Results of the Woodroffe Corridor

Table 13.46: Comparison of Mean Travel Times for the Car and Transit Modes – Base Case Scenario

Comparison of Travel Times for all Segments

		n = 20			Student's t distribution D.F.=20+20-2=38			
		Before	After	Diff	Pooled StDev	t	1.68 t(.95)	2.42 t(.99)
Car	TT	271	280	9				
	StDev	33	36		35.2	0.85	Not Sign	Not Sign
Bus	TT	228	213	-15				
	StDev	8	2		5.9	7.87	Sign	Sign
Car+Bus	TT	269	278	9				
	StDev	32	34		34.1	0.80	Not Sign	Not Sign

Comparison of Travel Times for Specific Segments

From To		Before	After	Diff				
1 A to C	TT	119	167	48				
	StDev	11	32		24.7	6.13	Sign	Sign
2 A to D	TT	181	227	46				
	StDev	8	30		22.6	6.45	Sign	Sign
3 A to E	TT	239	262	23				
	StDev	34	38		37.0	1.96	Sign	Not Sign
4 A to G	TT	293	298	5				
	StDev	45	42		45.0	0.37	Not Sign	Not Sign
5 A to H	TT	392	395	2				
	StDev	61	60		62.0	0.12	Not Sign	Not Sign
6 C to E	TT	64	54	-11				
	StDev	12	3		9.4	3.57	Sign	Sign
7 C to G	TT	109	84	-25				
	StDev	25	7		19.1	4.18	Sign	Sign
8 C to H	TT	240	201	-39				
	StDev	45	25		37.4	3.33	Sign	Sign
9 D to E	TT	202	158	-44				
	StDev	55	40		49.5	2.82	Sign	Sign
10 D to G	TT	263	223	-40				
	StDev	58	31		48.0	2.64	Sign	Sign
11 D to H	TT	417	354	-63				
	StDev	77	41		63.0	3.17	Sign	Sign
12 H to C	TT	147	136	-11				
	StDev	6	8		7.4	4.72	Sign	Sign
13 H to B	TT	130	133	3				
	StDev	13	18		16.0	0.51	Not Sign	Not Sign
14 H to A	TT	196	196	0				
	StDev	1	1		0.8	1.51	Not Sign	Not Sign
15 A to F	TT	270	246	-23				
	StDev	14	3		10.6	6.95	Sign	Sign
16 C to F	TT	155	135	-20				
	StDev	13	3		9.3	6.80	Sign	Sign
17 F to A	TT	216	219	3				
	StDev	2	2		2.2	4.43	Sign	Sign
18 F to C	TT	184	174	-10				
	StDev	7	8		7.6	4.22	Sign	Sign

Appendix T: Simulation Parameters for Fuzzy Logic NVDZ Estimation

This Appendix contains background data used in the development and testing of the queue length estimation method. The information is grouped into the following categories:

- summary of results;
- detection setup scenarios and related parameters;
- signal timing scenarios;
- volume scenarios;
- membership function parameters; and
- rule-base output values.

Note:

First line for each Data Set is for the training data and the second line for the testing data
Data Sets 32, 33, 34, and 35 were not used to generate FIS parameters, they were used only for testing.

Table 13.47: Summary of Results

Data Set	Error ML	R ²	m	RMS E	e	Max E	Timing Scenario	Detection Scenario	Volume Scenario	Comments
13	1.721	.9858 .987	0.993	1.87 1.83	1.28	10.3	50/15	A	A	Base case
10	1.77	.985 .9464	1.089	1.87 3.08	2.72	12.3	50/15	A(one lane)	A	Detectors in one lane
11	1.492	.983 .9768	0.856	1.57 3.19	2.63	12.4	50/15	A(one lane)	A	Similar to Set 10-swapped training and testing data
12	1.06	.981 .972	1.000	1.12 0.94	0.67	4.8	50/15	A	B	Traffic in one lane only
14	1.585	.979 .9832	1.007	1.86 1.86	1.30	10.7	50/15	A	A-First 7200 sec	Reduced data set to 7200 records
17	2.416	.9623 .965	0.91	2.46 2.98	2.31	11.7	50/15	B	C	Detection location moved towards the signal
19	2.27	.966 .964	0.924	2.35 2.87	2.22	14.0	50/15	B	C	Similar to Set 17 - Detection loop length reduced from 6 to 5 m
18	2.49	.9468 .9337	1.054	2.93 3.73	3.16	13.8	50/15	C	C	Detection location moved further towards the signal
20	1.52	.973 .974	0.998	1.76 1.77	1.3	8.2	50/15	D	A	Reduced detection zone from 180 to 120 m
21	2.53	.976 .985	1.05	3.09	2.39	15.5	50/15	E	A	Increased detection zone from 180 to 240 m
30	2.7	.905 .917	1.052	2.86	2.31	9.66	130/60	A	D	Signal timing changed for longer cycle and stage times
32	n/a	.885	.776	6.16	4.94	21.7	130/60 (50/15)	A	A	ANFIS Parameters for 130/60 (Data Set 30) used for Signal Timing 50/15
33	n/a	.89	.937	3.84	2.85	17.5	50/15 (130/60)	A	D	ANFIS Parameters for 50/15 (Data Set 13) used for Signal Timing 130/60
34	n/a	.859	1.016	3.69	2.74	14.9	50/15 (50/25)	A	D	ANFIS Parameters for 50/15 (Data Set 13) used for Signal Timing 50/25
35	n/a	.979	1.078	3.05	2.31	10.9	50/15 (50/20)	A	E	ANFIS Parameters for 50/15 (Data Set 13) used for Signal Timing 50/15

Error ML – error reported by MATLAB’s ANFIS

R^2 – coefficient of determination. Since the assumed relationship is linear, the correlation coefficient is the square root of R^2

m – tangent of the least square line

RMSE – root mean square

$|e|$ - mean absolute error

Max $|E|$ - maximum absolute error

Table 13.48: Detector Locations for Various Scenarios

	Scenario A (m)	Scenario B (m)	Scenario C (m)	Scenario D (m)	Scenario E (m)
L1	30	30	30	20	30
L2	60	60	60	40	90
L3	30	30	30	20	30
L4	180	180	180	120	240
L5	240	5	-5	240	240

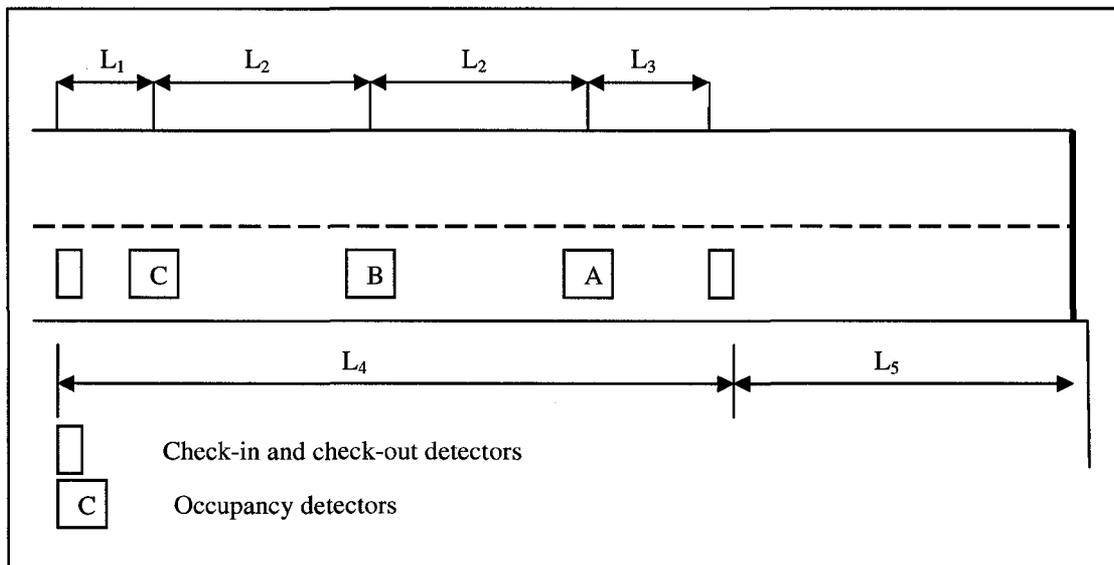


Figure 13.27: Detector Set up

Table 13.49: Detector Parameters in VISSIM

Detector Type	Length (m)	Smoothing Factors (Increase/Decrease)
Counting	1	n/a
Occupancy	6	0.05/0.05

Table 13.50: Signal Timing Scenarios

	Scenario A (sec)	Scenario B (sec)	Scenario C (sec)	Scenario D (sec)
Cycle	50			130
Green Stage	15	20	25	60

Table 13.51: Volume Scenarios

Time Period (sec)	Scenario A		Scenario B		Scenario C		Scenario D		Scenario E
	Train. (veh/ hr)	Test (veh/ hr)	Train. (veh/ hr)	Test (veh/ hr)	Train. (veh/ hr)	Test (veh/ hr)	Train. (veh/ hr)	Test (veh/ hr)	Test (veh/ hr)
0-900	1100	200	550	150	800	300	1700	1100	800
900-1800	200	1100	100	250	300	800	1100	1700	1300
1800-2700	1100	200	550	550	800	300	1700	1100	800
2700-3600	200	1100	100	150	300	800	1100	1700	1300
3600-4500	1100	200	550	250	800	300	1700	1100	800
4500-5400	200	1100	100	550	300	800	500	500	1300
5400-6300	1100	200	550	150	800	300	1700	1100	800
6300-7200	200	1100	100	250	300	800	1100	1700	1300
7200-8100	1100	200	550	550	800	300	1700	1100	800
8100-9000	200	1100	100	150	300	800	1100	1700	1300
9000-9900	1100	200	550	250	800	300	1700	1100	800
9900-10800	200	1100	100	550	300	800	500	500	1300
18000-11700	1100	200	550	150	800	300	1700	1100	800
11700-12600	200	1100	100	250	300	800	1100	1700	1300
12600-13500	1100	200	550	550	800	300	1700	1100	800
13500-14400	200	1100	100	150	300	800	1100	1700	1300

Table 13.52: Membership Function Parameters

		Data Set											
		10	11	12	13	14	17	18	19	20	21	30	31
OccA_S	1	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99
	2	-95	-95	-95	-95	-95	-95	-95	-95	-95	-95	-95	-95
	3	.154	.193	.149	.177	.152	.15	.13	.16	.14	.15	.17	.15
	4	.371	.439	.362	.404	.355	.35	.34	.38	.35	.36	.460	.35
OccA_M	1	.155	.125	.18	.172	.148	.14	.078	.15	.14	.15	.13	.14
	2	.352	.381	.346	.377	.349	.35	.33	.36	.34	.36	.36	.34
	3	.705	.673	.645	.749	.710	.69	.81	.76	.72	.68	.83	.85
	4	.989	.964	.99	.991	.99	.99	.99	.98	.99	.99	.99	.94
OccA_L	1	.598	.550	.578	.569	.664	.6	.71	.55	.67	.62	.50	.49
	2	.881	.848	.831	.891	.899	.87	.98	.91	.91	.86	.85	.81
	3	95	95	95	95	95	95	95	95	95	95	95	95
	4	99	99	99	99	99	99	99	99	99	99	99	99
OccB_S	1	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99
	2	-95	-95	-95	-95	-95	-95	-95	-95	-95	-95	-95	-95
	3	.04	.364	.112	.115	.144	.14	.07	.17	.13	.12	.15	.15
	4	.384	.367	.348	.442	.395	.44	.41	.45	.34	.42	.41	.40
OccB_M	1	.010	.039	.019	.03	.042	.03	.07	.024	.05	.02	.11	.11
	2	.198	.206	.302	.278	.326	.32	.25	.30	.33	.30	.35	.34
	3	.649	.649	.649	.656	.64	.66	.74	.67	.64	.71	.65	.69
	4	.891	.874	.875	.978	.882	.89	.99	.94	.88	.99	.81	.98
OccB_L	1	.639	.622	.619	.573	.588	.64	.72	.65	.59	.61	.66	.59
	2	.846	.846	.846	.84	.835	.85	.94	.87	.84	.87	.84	.81
	3	95	95	95	95	95	95	95	95	95	95	95	95
	4	99	99	99	99	99	99	99	99	99	99	99	99
OccC_S	1	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99
	2	-95	-95	-95	-95	-95	-95	-95	-95	-95	-95	-95	-95
	3	.046	.036	.053	.119	.117	.13	.11	.098	.11	.12	.18	.18
	4	.345	.439	.408	.409	.402	.41	.36	.41	.41	.44	.56	.54
OccC_M	1	.006	.022	.011	.024	.03	.02	.01	.017	.04	.01	.11	.11
	2	.222	.212	.276	.328	.328	.32	.32	.28	.31	.30	.37	.36
	3	.696	.695	.69	.687	.647	.66	.68	.70	.68	.64	.63	.63
	4	.986	.923	.992	.990	.896	.92	.98	.99	.99	.88	.87	.86
OccC_L	1	.623	.679	.656	.605	.603	.64	.63	.66	.52	.61	.51	.56
	2	.872	.896	.884	.872	.843	.86	.88	.89	.85	.84	.82	.83
	3	95	95	95	95	95	95	95	95	95	95	95	95
	4	99	99	99	99	99	99	99	99	99	99	99	99

Table 13.53: Rule Output Values

Rule		Data Set											
		10	11	12	13	14	17	18	19	20	21	30	31
1	SSS	1.35	.169	.073	.06	.015	0.58	1.56	0.59	.33	.19	5.9	5.8
2	SSM	11.5	6.94	10.9	12.5	12.3	16.7	22.3	15.7	9.9	16.12	17.9	17.8
3	SSL	0	0	0	0	0	0	0	0	0	0	0	0
4	SMS	2.69	0	0	0	0	1.4	8.2	2.1	0	0	0	0
5	SMM	12.7	12.8	10.4	10.1	14	9.4	0	7.1	13.9	16.1	7.2	11.7
6	SML	0	0	0	0	0	0	0	0	0	0	0	0
7	SLS	0	0	0	0	0	0	0	0	0	0	0	0
8	SLM	0	0	0	0	0	0	0	0	0	0	0	0
9	SLL	0	0	0	0	0	0	0	0	0	0	0	0
10	MSS	3.36	2.14	1.48	2.24	1.78	1.9	2.98	2.7	1.0	4.3	5.4	5.7
11	MSM	11.1	20.2	10.2	5.26	0.8	12.8	22.3	11.0	1.4	5.9	13.8	10.7
12	MSL	0	0	0	0	0	0	0	0	0	0	0	0
13	MMS	12.9	6.66	2.03	2.96	3.2	3.5	7.3	5.4	0.86	7.2	6.5	4.3
14	MMM	14.9	14.2	9.20	10.5	11.7	15.2	12.7	16.1	6.7	17.1	17.7	16.8
15	MML	0	0	0	0	0	0	0.29	20	0	0	18.9	9.6
16	MLS	16.9	20.6	0	0	0	0	0.43	6	0	13	36.9	24.8
17	MLM	20.9	6.41	14.5	0	0	0	34	11.2	21	15	10.2	12.9
18	MLL	37.2	32.6	11.6	0	0.13	0	33	6.9	27.6	22	24.3	25.8
19	LSS	9.67	6.98	4.03	7.3	7.5	8.6	11.6	9.1	3.4	13	14.8	6.9
20	LSM	17.3	17.8	14.1	11.3	11.3	15.2	5.9	16.4	12.4	16.3	3	22.5
21	LSL	0	0	0	0	0	0	0	0	0	0	0	0
22	LMS	19.9	15.5	9.13	16.9	17.2	20	18.7	21.1	10.2	25.5	11.4	13.2
23	LMM	29.3	26.1	14.4	21.6	21.0	21	26.3	23	13.4	30.5	23.7	21.6
24	LML	0	0	0	30	0	0	0.37	45.1	39.6	40	31	30.1
25	LLS	22.1	20.4	10.8	19.9	19.3	21.2	24.3	22.5	12.0	34.7	6.3	9.4
26	LLM	34.3	33.2	16.9	30.7	28.9	32.7	33.4	34.5	19.7	48.2	27.4	29
27	LLL	42.3	38.5	20.6	40.5	39.8	40.2	42.8	42.1	26.2	58.8	31.1	32.7

Table 13.54: Comparison of Output Results Based on Vehicle Counts and Fuzzy Logic NVDZ Estimation

Comparison of Travel Times for all Segments

		n			Student's t distribution D.F.=20+20-2=38			
		Count	FL Estimate	Diff	Pooled StDev	t	1.68 t(.95)	2.42 t(.99)
Car	TT	280.0	280.8	0.8				
	StDev	35.6	35.7		36.6	0.07	Not Sign	Not Sign
Bus	TT	213.2	212.9	-0.3				
	StDev	1.7	1.8		1.8	0.61	Not Sign	Not Sign
Car+Bus	TT	277.6	278.4	0.7				
	StDev	34.3	34.42		35.3	0.07	Not Sign	Not Sign

Comparison of Travel Times for Specific Segments

	From To	Count	FL Estimate	Diff				
1	A to C	TT	167	172	5			
		StDev	32	35		34.4	0.46	Not Sign
2	A to D	TT	227	232	5			
		StDev	30	32		31.6	0.47	Not Sign
3	A to E	TT	262	268	6			
		StDev	38	40		40.2	0.50	Not Sign
4	A to G	TT	298	297	-1			
		StDev	42	41		42.8	0.06	Not Sign
5	A to H	TT	395	394	-1			
		StDev	60	59		61.1	0.04	Not Sign
6	C to E	TT	54	54	0			
		StDev	3	3		3.5	0.01	Not Sign
7	C to G	TT	84	83	-1			
		StDev	7	6		6.5	0.59	Not Sign
8	C to H	TT	201	194	-7			
		StDev	25	23		24.2	0.86	Not Sign
9	D to E	TT	158	156	-3			
		StDev	40	41		41.9	0.19	Not Sign
10	D to G	TT	223	215	-8			
		StDev	31	32		32.8	0.73	Not Sign
11	D to H	TT	354	346	-7			
		StDev	41	42		42.6	0.53	Not Sign
12	H to C	TT	136	136	-1			
		StDev	8	9		8.8	0.24	Not Sign
13	H to B	TT	133	137	5			
		StDev	18	13		16.3	0.88	Not Sign
14	H to A	TT	196	196	0			
		StDev	1	1		1.0	0.03	Not Sign
15	A to F	TT	246	246	-1			
		StDev	3	3		3.2	0.83	Not Sign
16	C to F	TT	135	134	-1			
		StDev	3	3		3.2	1.09	Not Sign
17	F to A	TT	219	220	1			
		StDev	2	3		2.8	0.87	Not Sign
18	F to C	TT	174	174	0			
		StDev	8	8		8.2	0.04	Not Sign

1