STRATEGIES FOR RAPID CONGESTION RECOVERY USING RAMP METERING

Rui JIANG
Bachelor of Engineering (Telecommunication)
Master of Engineering (Telecommunication)

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Smart Transport Research Centre
Civil Engineering and Built Environment School
Science and Engineering Faculty
Queensland University of Technology

2014
Keywords

Motorway congestion, Rapid congestion recovery, Ramp metering, Cell transmission model, On-ramp queue estimation, Queue management, Ramp coordination, Micro-simulation.
Abstract

Congestion is one of the biggest problems for urban motorway mobility worldwide. In Australia particularly, urban motorways carry the commuting traffic in peak hours and most of the freight traffic. These long periods of motorway congestion, caused by high demand, are creating major problems for road users and the community.

Many motorway management tools for tackling congestion, such as ramp metering (RM) and variable speed limits (VSL), have been developed to reduce traffic congestion in metropolitan motorway networks. Among current available motorway management tools, RM is considered to be the most effective tool for motorway congestion, with worldwide utilisation, and its effectiveness has already been proven by field applications. However, motorway congestion can be only reduced, never eliminated, by RM, due to the excessively increased long peak hours and such field constraints of RM operation as limited ramp storage space and maximum ramp waiting time. This research, therefore, develops a post-congestion strategy, that is rapid congestion recovery (RCR), using RM.

The research can be divided into two stages. In the first stage, a coordinated RM system was built up as the bench mark, since the existing literature documents that this type of RM can well represent current practice. Two major components for a coordinated RM system were studied. The first one was a sophisticated local queue management algorithm: a real-time on-ramp queue estimation algorithm was developed based on Kalman filter theory to provide accurate queue information for the management algorithm. The second one was a coordination strategy between on-ramps. The coordinated RM algorithm adopts a two-layer control structure. At the higher layer, with long update interval, ramp coordination is planned and arranged: that is to assemble / disassemble coordination group based on the location of high-risk breakdown flow. At the lower layer, with short update interval, the coordination is executed dynamically: PID (Proportion-Integration-Differentiation) controllers are developed to control individual ramps hired to serve the coordination, and dynamic release of hired ramp is based on the prevailing congestion level on the ramp. These two components together created the bench mark.
In the second stage, this research concentrated on RCR using RM. Firstly, the recovery phase was defined and traffic flow dynamics at recovery phase were discussed. By doing so, the potential benefits of RCR were well analysed. According to the analyses, restrictive metering control (RMC), which operates the most restrictive metering rate, was selected as the basic operation for recovery, and a micro-simulation investigation demonstrated that RMC is able to improve the merging bottleneck throughput at the recovery phase. Secondly, a RM strategy for RCR, called the zone-based RM strategy for rapid congestion recovery (ZRM-RCR), was developed by dividing the motorway network into zones based on mainline queues. In ZRM-RCR, a two-phase control algorithm (the compulsory control phase and the reactive control phase) is proposed. In the compulsory control phase, bottleneck ramps are forced to activate RMC for rapid mainline queue discharge; in the reactive control phase, ramp costs (including queue spillover and ramp queues) are considered and managed. This strategy was tested in two test-beds in a micro-simulation environment for recurrent congestion scenario. Evaluation results justified the effectiveness of ZRM-RCR: that is, the strategy can accelerate the system recovery and manage the total ramp costs at the same time. Thirdly, minor modifications were then made in ZRM-RCR for incident scenario, and evaluation results also implied its benefits.

As all the previous results rely on the micro-simulation model, the ZRM-RCR was then evaluated in a cell transmission model (CTM) at a macroscopic level. In order to model metered merging more precisely, this research proposed a hybrid modelling framework that would build the numerical relationship between the merging capacity and the metering rate. This enables variable merging capacity in merge cell of CTM, and the modified model is called modified CTM (M-CTM). The evaluation results from the M-CTM further confirmed the effectiveness and benefits of ZRM-RCR.
# Table of Contents

Keywords .................................................................................................................................................. i
Abstract .................................................................................................................................................. iii
Table of Contents .................................................................................................................................. v
List of Figures .......................................................................................................................................... ix
List of Tables .......................................................................................................................................... xi
List of Abbreviations .............................................................................................................................. xiii
Statement of Original Authorship ......................................................................................................... xv
Acknowledgements ............................................................................................................................... xvii

## CHAPTER 1 INTRODUCTION .............................................................................................................. 1
1.1 Background and Motivation ............................................................................................................. 1
1.2 Research Problem ........................................................................................................................... 3
  1.2.1 Problem Statement .................................................................................................................... 3
  1.2.2 Research Questions ................................................................................................................... 4
1.3 Aims .................................................................................................................................................. 5
1.4 Contributions ................................................................................................................................... 6
1.5 Thesis Outline .................................................................................................................................. 7

## CHAPTER 2 LITERATURE REVIEW .................................................................................................... 9
2.1 Introduction ....................................................................................................................................... 9
2.2 Evolutionary Progress in Ramp Metering Systems ........................................................................ 9
  2.2.1 Efforts for Accurate Information in Real Time ......................................................................... 9
  2.2.2 Efforts for Sophisticated Control Algorithms .......................................................................... 11
  2.2.3 Efforts for Effective Actuators .................................................................................................. 12
2.3 Evolution of Ramp Metering ........................................................................................................ 13
  2.3.1 Pre-determined RM Systems .................................................................................................... 13
  2.3.2 Local Responsive RM Systems ............................................................................................... 14
  2.3.3 Coordinated RM Systems ....................................................................................................... 17
  2.3.4 RM integrated in ITS .............................................................................................................. 22
2.4 RM Systems in Practice ................................................................................................................. 24
  2.4.1 State of Practice ...................................................................................................................... 24
  2.4.2 Benefits of RM Implementation .............................................................................................. 25
  2.4.3 Issues Raised by Field Implementation ................................................................................... 26
  2.4.4 Lessons Learned from Practice ............................................................................................. 27
2.5 Internet Congestion Control ......................................................................................................... 28
  2.5.1 Why Internet Congestion Control .......................................................................................... 28
  2.5.2 Overview .................................................................................................................................. 28
  2.5.3 TCP Congestion Control .......................................................................................................... 29
2.6 Summary ......................................................................................................................................... 32

## CHAPTER 3 LOCAL QUEUE MANAGEMENT .................................................................................... 33
3.1 Introduction ..................................................................................................................................... 33
3.2 Framework ..................................................................................................................................... 34
3.3 Queue Estimation .......................................................................................................................... 34
  3.3.1 Existing techniques for Ramp Queue Estimation .................................................................... 35
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.3.1</td>
<td>Compulsory Control</td>
<td>104</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Reactive Control</td>
<td>105</td>
</tr>
<tr>
<td>6.4</td>
<td>Evaluation of Peak Hour Congestion Scenario</td>
<td>108</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Simulation Settings</td>
<td>108</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Performance Indicators</td>
<td>109</td>
</tr>
<tr>
<td>6.4.3</td>
<td>Parameter Calibration</td>
<td>109</td>
</tr>
<tr>
<td>6.4.4</td>
<td>Results from Test-bed 1</td>
<td>111</td>
</tr>
<tr>
<td>6.4.5</td>
<td>Results from Test-bed 2</td>
<td>115</td>
</tr>
<tr>
<td>6.4.6</td>
<td>Macroscopic Fundamental Diagram Analysis for Test-bed 2</td>
<td>120</td>
</tr>
<tr>
<td>6.5</td>
<td>Evaluation of Incident Scenario</td>
<td>125</td>
</tr>
<tr>
<td>6.5.1</td>
<td>Incident Settings</td>
<td>125</td>
</tr>
<tr>
<td>6.5.2</td>
<td>Recovery Phase Identification for Incident Scenario</td>
<td>125</td>
</tr>
<tr>
<td>6.5.3</td>
<td>Simulation Settings</td>
<td>126</td>
</tr>
<tr>
<td>6.5.4</td>
<td>Result Analysis</td>
<td>127</td>
</tr>
<tr>
<td>6.6</td>
<td>Summary</td>
<td>130</td>
</tr>
</tbody>
</table>

**CHAPTER 7**  EVALUATION USING A MODIFIED CELL TRANSMISSION MODEL......131

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1</td>
<td>Introduction</td>
<td>131</td>
</tr>
<tr>
<td>7.2</td>
<td>Merge Cell in CTM and its Weakness</td>
<td>132</td>
</tr>
<tr>
<td>7.3</td>
<td>Variable Merging Capacity Relating to Metering Rate</td>
<td>135</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Moving Bottleneck Theory</td>
<td>135</td>
</tr>
<tr>
<td>7.3.2</td>
<td>Linear Car-Flowing Model</td>
<td>136</td>
</tr>
<tr>
<td>7.3.3</td>
<td>Hybrid Modelling Framework</td>
<td>137</td>
</tr>
<tr>
<td>7.3.4</td>
<td>Model Settings and Results</td>
<td>138</td>
</tr>
<tr>
<td>7.4</td>
<td>Evaluation</td>
<td>141</td>
</tr>
<tr>
<td>7.4.1</td>
<td>Simulation Network in CTM</td>
<td>141</td>
</tr>
<tr>
<td>7.4.2</td>
<td>Modelling RM in CTM</td>
<td>142</td>
</tr>
<tr>
<td>7.4.3</td>
<td>Performance Indicators</td>
<td>144</td>
</tr>
<tr>
<td>7.4.4</td>
<td>Results</td>
<td>144</td>
</tr>
<tr>
<td>7.5</td>
<td>Summary</td>
<td>146</td>
</tr>
</tbody>
</table>

**CHAPTER 8**  DISCUSSION AND CONCLUSION.................................................147

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1</td>
<td>Response to Research Questions</td>
<td>147</td>
</tr>
<tr>
<td>8.2</td>
<td>Limitations and Future Research</td>
<td>149</td>
</tr>
<tr>
<td>8.3</td>
<td>Concluding Remarks</td>
<td>150</td>
</tr>
<tr>
<td>REFERENCES</td>
<td></td>
<td>153</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1-1 Conceptual cumulative volume curve ................................................................. 4
Figure 2-1 Control loop of RM system .................................................................................... 10
Figure 2-2 Evolution of RM .................................................................................................. 13
Figure 2-3 Model predictive control structure (Papamichail, Kotsialos, Margonis, & Papageorgiou, 2008) .......................................................... 21
Figure 3-1 Local RM sub-system ......................................................................................... 34
Figure 3-2 Metered motorway on-ramp and detector requirements ...................................... 37
Figure 3-3 Estimation RMSE in dependence of the Kalman gain K ..................................... 43
Figure 3-4 Mid-link time occupancies and number of vehicles ........................................... 44
Figure 3-5 Queue estimation algorithm flow ................................................................. 45
Figure 3-6 Road geometry of the selected on-ramps for the algorithm evaluation .............. 46
Figure 3-7 Traffic flow profiles (15 minutes period) ............................................................. 47
Figure 3-8 Queue estimation results example (Birdwood Road ramp) ................................. 50
Figure 3-9 Queue estimation results example (Marquis Street ramp) .................................. 51
Figure 3-10 Queue estimation results example (Logan Road ramp) ................................... 51
Figure 3-11 Local RM rate determination process ............................................................ 56
Figure 3-12 Metering control example – QM scheme vs. QF strategy at Birdwood Road ramp ....61
Figure 4-1 Multilayer control framework ........................................................................... 64
Figure 4-2 Flowchart of CRM algorithm ........................................................................... 66
Figure 4-3 Smoothed data for coordination activation....................................................... 67
Figure 4-4 Queue projection .............................................................................................. 68
Figure 4-5 Master requirement ........................................................................................ 70
Figure 4-6 Slave contribution ........................................................................................... 71
Figure 4-7 The test-bed - Northbound of the Pacific Motorway ......................................... 78
Figure 4-8 Motorway mainline speed contour (base case vs. LRM) ..................................... 82
Figure 4-9 Motorway mainline speed contour (LRM vs. CRM) ......................................... 83
Figure 5-1 Speed contour of the new simulation model - Modified northbound of the Pacific Motorway ........................................................................... 90
Figure 5-2 Detector settings .............................................................................................. 91
Figure 5-3 Speed contour of Test-bed L with 20-minute RMC ........................................... 93
Figure 6-1 Definition of zone ............................................................................................ 98
Figure 6-2 Zone formulation ............................................................................................ 98
Figure 6-3 Framework of ZRM-RCR .................................................................................. 99
Figure 6-4 Detection of mainline queue tail ..................................................................... 100
Figure 6-5 Mainline queue movement at recovery phase .................................................... 101
Figure 6-6 Flow chart of recovery phase identification ....................................................... 102
Figure 6-7 Two-phase control structure for ZRM-RCR ..................................................... 103
List of Tables

Table 2-1 List of rule-based heuristic algorithms..........................................................19
Table 2-2 List of model-based system optimization algorithms ......................................20
Table 3-1 Simulation results of the space occupancy estimation and the measured queue length ........42
Table 3-2 Queue Estimation Algorithm Evaluation Results ............................................49
Table 3-3 Queue management evaluation – Birdwood Road ramp ..................................58
Table 3-4 Queue management evaluation – Marquis Street ramp ..................................59
Table 3-5 Queue management evaluation – Logan Road ramp ....................................60
Table 4-1 Evaluation results of 2 aggregation periods ..................................................69
Table 4-2 Evaluation results for parameter calibration ..................................................69
Table 4-3 PID terms in RM .........................................................................................74
Table 4-4 Analysis of the controller term effectiveness ..................................................75
Table 4-5 Grouping slave ramps ..................................................................................76
Table 4-6 Candidate of controller structure ..................................................................77
Table 4-7 PID controller formulation ...........................................................................77
Table 4-8 Coordinated ramp metering evaluation results summary ..................................80
Table 4-9 Comparison of average ramp traffic delay by individual ramps ......................82
Table 5-1 20-minute throughput results of Test-bed L ...................................................92
Table 5-2 20-minute average speed results of Test-bed L .............................................92
Table 5-3 20-minute throughput results of Test-bed S ..................................................94
Table 5-4 20-minute average speed results of Test-bed S .............................................94
Table 6-1 Calibration results of consecutive interval for recovery phase identification .........110
Table 6-2 Calibration results of compulsory control length from Test-bed 1 ....................110
Table 6-3 Calibration results of compulsory control length from Test-bed 2 .................111
Table 6-4 Simulation result summary of Test-bed 1 ......................................................111
Table 6-5 Individual ramp queue spillover results of Test-bed 1 ....................................112
Table 6-6 Individual ramp traffic travel time in the recovery period of Test-bed 1 ..........113
Table 6-7 Simulation result summary of Test-bed 2 ......................................................116
Table 6-8 Comparison of the performance indicators (CRM vs. ZRM-RCR) .................118
Table 6-9 Individual ramp queue spillover time comparison (CRM vs. ZRM-RCR) ...........119
Table 6-10 Individual average ramp traffic travel time at the recovery period comparison (CRM vs. ZRM-RCR) .................................................................120
Table 6-11 Simulation result summary of the incident scenario .....................................127
Table 7-1 Model parameters .......................................................................................138
Table 7-2 Simulation results summary ..........................................................................139
Table 7-3 Simulation results summary using modified CTM ........................................145
Table 7-4 Individual ramp results using modified CTM ...............................................145
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIMR</td>
<td>Additive Increase Minimal Release algorithm</td>
</tr>
<tr>
<td>CC-AIMR</td>
<td>Cost Constrained-AIMR algorithm</td>
</tr>
<tr>
<td>CF(L)</td>
<td>Linear Car Following model</td>
</tr>
<tr>
<td>CTM</td>
<td>Cell Transmission Model</td>
</tr>
<tr>
<td>CRM</td>
<td>Coordinated Ramp Metering</td>
</tr>
<tr>
<td>ITS</td>
<td>Intelligent Transportation Systems</td>
</tr>
<tr>
<td>KW(T)</td>
<td>Kinematic Wave model with Triangular fundamental diagram</td>
</tr>
<tr>
<td>LRM</td>
<td>Local Ramp Metering</td>
</tr>
<tr>
<td>MAE</td>
<td>Mean Average Error</td>
</tr>
<tr>
<td>M-CTM</td>
<td>Modified CTM</td>
</tr>
<tr>
<td>MFD</td>
<td>Macroscopic Fundamental Diagram</td>
</tr>
<tr>
<td>MPE</td>
<td>Mean Percentage Error</td>
</tr>
<tr>
<td>MTD</td>
<td>Mainline Traffic Delay</td>
</tr>
<tr>
<td>NV</td>
<td>Number of Vehicles</td>
</tr>
<tr>
<td>PID</td>
<td>Proportion Integration Differentiation controller</td>
</tr>
<tr>
<td>QF</td>
<td>Queue Flushing</td>
</tr>
<tr>
<td>QM</td>
<td>Queue Management</td>
</tr>
<tr>
<td>QSOT</td>
<td>Queue SpillOver Time</td>
</tr>
<tr>
<td>RCR</td>
<td>Rapid Congestion Recovery</td>
</tr>
<tr>
<td>RM</td>
<td>Ramp Metering</td>
</tr>
<tr>
<td>RMC</td>
<td>Restrictive Metering Control</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>RTD</td>
<td>Ramp Traffic Delay</td>
</tr>
<tr>
<td>TQST</td>
<td>Total Queue Spillover Time</td>
</tr>
<tr>
<td>TTT</td>
<td>Total Travel Time</td>
</tr>
<tr>
<td>VSL</td>
<td>Variable Speed Limits</td>
</tr>
<tr>
<td>ZRM-RCR</td>
<td>Zone-based Ramp Metering strategy for RCR.</td>
</tr>
</tbody>
</table>
Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signature:

Date: 28/02/2014
Acknowledgements

My PhD research was completed under the sponsorship of the Smart Transport Research Centre (STRC). I would like to take this opportunity to acknowledge the substantial financial assistance that I received.

I have complex and rich feelings about this PhD journey. During the entire life of this journey, there have been struggles and passions, frustrations and achievements, stress and pleasure, suffering and benefits, loneliness and teamwork, as well as many other feelings. However, I was not alone on this journey. I felt how lucky I was to have so many people around me, supervising me, helping me, supporting me, and encouraging me. I am deeply grateful to these people.

First and foremost, I would like to extend my utmost gratitude to my supervisors, Prof. Edward Chung and Associate Prof. Jonathan Bunker, who introduced me to this body of scholarship, guided me through this project, and encouraged me to be the best that I can. And especially for my principal supervisor, Prof. Edward Chung, he was so engaged in my project that he contributed significant time to supervise me, without which I would never have reached the end of this PhD journey.

Secondly, I am indebted to my beloved wife, Yuan Liu; to my parents, Jianping Jiang and Xuelan Feng; and to my parents-in-law, Jiangping Liu and Yu Ruan, who have all accompanied me through the ups and downs along my PhD journey, helped me build up my confidence, and encouraged me never to give up. I hope I have not failed their love and expectations.

Thirdly, special thanks are also given to Dr. Brian Lee, who has also been very involved in the research and who has helped me with many technical issues.

Fourthly, I would like to express my sincere gratitude to the STRC term. Their constructive comments and thought-provoking advice helped me conquer every technical obstruction in my PhD journey.

Last but not least, I would like to extend my thanks, from the depth of my heart, to all my friends who supported me during the entire three years of this PhD, both in academic and in daily life: Yaxing Zhang, Yang Hu, Jianwei Liu, Xiangguo
Wang, Juncheng Dai, Ark Du, Yuming Guo, Jun Wang, Yi Yu, Ruizi Wang, Fengfeng Li, Wenbo Ma and Nannan Zong. Without the friendship of these people, my PhD journey would have been a lonely one.

I owe thanks to all these people, who witnessed my struggle and who cheered me on. Without the help, support, encouragement, concern, trust, and love of these people, I would not have been able to complete this thesis. This fulfilment does not belong to me alone: it belongs to all these beloved people. By presentation of this thesis, I would like to show them how this PhD journey has become a miracle in my life!
Chapter 1 Introduction

1.1 BACKGROUND AND MOTIVATION

Motorways are designed to provide high levels of mobility to road users. However, the dramatic expansion of cities and the continuous growth of car ownership have led to heavy congestion, which strongly reduces traffic throughput, fluidity and safety, as well as increasing trip times and environmental pollution. Motorway congestion has become a worldwide problem for both the motorway administration and the road users.

Population increases in Australia have resulted in a significant growth in traffic demand. A newspaper report (Moore, 2010) indicates that morning peak hour is now from 5:00 am on some Brisbane motorways, such as the Ipswich Motorway and the Pacific Motorway, which is two hours earlier than in 2002. This is because 1) urban motorways carry most of the commuting traffic in peak hours; and 2) the majority of motorways are now operating without any control strategies. The long periods of heavy congestion on motorways are becoming an annoyance for road users and communities. According to the Australian Government Department of Transport and Regional Services (2007), the cost of congestion in Australia, estimated at around $9.4 billion in 2005, is expected to rise to over $20.4 billion by 2020.

In order to tackle motorway congestion, many motorway management tools, such as ramp metering (RM) and variable speed limits (VSL), have been developed to reduce traffic congestion in metropolitan motorway networks. RM is considered to be the most effective tool currently available for motorway congestion, with its effectiveness already proven by field implementation results (M Papageorgiou & Kotsialos, 2002). In a metered on-ramp, a traffic signal is placed to regulate the rate of vehicles entering motorways. Normally, the metering rate is determined by real-time system conditions for both mainstream and ramp.

Since the first use of RM on the Eisenhower Expressway (I-290) in Chicago, Illinois in 1963 (Pinnell, Drew, McCasland, & Wattleworth, 1967; Wattleworth & Berry, 1965), many RM strategies have been proposed and some have been applied in the field. Field evaluation results from the literature (Muhurdarevic et al., 2006;
Piotrowicz & Robinson, 1995) indicate that RM systems are successful in delaying the onset of congestion and reducing congestion. However, they cannot eliminate congestion, even with the latest RM strategy (Papamichail, Papageorgiou, Vong, & Gaffney, 2010), because of the conflict between RM’s control objective and the field circumstances in operation. The major objective of these strategies is to prevent congestion and to maintain free flow on the motorway. As long as there is enough space in on-ramps for holding ramp traffic, this objective can be achieved. However, field limitations and expanded peak hours with high traffic demand make this impossible in practice.

On the one hand, there are some field constraints for RM system, such as maximum ramp waiting time and limited ramp storage space. With short time holding ramp traffic, both mainline traffic and ramp traffic can benefit if no congestion happens. Even for ramp traffic, free flow conditions on mainline provide them better opportunities to enter and use motorways. However, long time queuing for ramp traffic is unacceptable and inequitable. More importantly, limited ramp storage space in reality would cause ramp queue spillover back to the upstream arterial roads, which could seriously impact upon surface traffic. Consequently, RM systems in the field must increase metering rates at certain points to limit ramp traffic waiting time and to reduce queue spillover. This operation is usually against the objective of keeping the mainline traffic flowing freely, thereby causing flow breakdown.

On the other hand, the expansion of peak hours accelerates the activation of these field limitations. As noted, morning peak hours in some motorways of Brisbane have been brought forward two hours, compared with a decade ago (Moore, 2010). It is impossible to prevent congestion by holding ramp traffic for such a long period.

From the above discussion, a conclusion is that RM can only delay and reduce motorway congestion, but not avoid it, given current field conditions. This suggests that a strategy for dealing with periods after congestion is worth researching. Additionally, to the best knowledge of the author, there is no previous study systematically investigating a strategy for motorway congestion recovery after breakdown. This research project, therefore, focuses on post-congestion strategies targeting rapid recovery.
1.2 RESEARCH PROBLEM

1.2.1 Problem Statement

RM uses mainline conditions to regulate ramp traffic entering motorways. This offers two main benefits: one is smooth merging behaviours by breaking large platoons; the other is mainline traffic in free flow condition by delaying ramp traffic. The main disadvantages of RM include ramp traffic delays and queue spillover back to adjacent arterial roads. Obviously, ramp traffic delays and queue spillover can be seen as the costs of RM. When the total demands from both mainline and ramps do not exceed capacity for a long time, a nearly free flow condition can be maintained in mainline without significant ramp delays. In addition, a smooth mainline traffic condition eventually benefits ramp traffic by increasing the opportunity to use the motorway. Overall, reasonable costs endured by ramp traffic improve system performance, and therefore reward all users in this situation.

Motorway congestion is unavoidable, however, due to excessively increased traffic demands during the expanded peak hours, as analysed in Section 1.1. This means that maintaining free flow mainline would require huge delays for ramp traffic and might affect arterial roads significantly with queue spillover. The costs are unacceptable; therefore, RM cannot stop motorway congestion. Once congestion happens, the benefits of RM are reduced.

Naturally, traffic demands reduce when peak hours end, providing another opportunity for RM to benefit the whole system, again with reasonable costs. As the mainline demand is decreasing, there would be no more mainline queue accumulation. Consequently, the earlier the mainline queue is cleared, the more travel time saving can be achieved. As the ramp demands are reducing, total ramp traffic delays might be able to be managed in an acceptable manner, and the risk of queue spillover is much smaller. Moreover, the nature of RM would give priority to mainline traffic, so RM is an appropriate tool for the purpose of rapid congestion recovery.

Taking all these into consideration led to the scope of this research project: to investigate RM strategies for post-congestion treatment – that is, rapid congestion recovery for motorways.
1.2.2 Research Questions

This sub-section firstly presents the key notions in this research problem, and then identifies the research questions. In order to accurately clarify the research questions, three key notions are stated as follows:

- Motorway mainline congestion: the state when the motorway mainline experiences traffic flow of low speed and high density, usually accompanied with mainline vehicle queuing. This state, also called “motorway mainline breakdown”, is caused by excess demand over capacity. After breakdown, capacity is usually observed to be lower than in the free flow condition (Srivastava & Geroliminis, 2013), which essentially increases delay in the motorway network.

- Recovery phase: the phase during which a motorway system recovers after mainline congestion, such as reduced demand at the end of peak hours or the clearance of a serious incident.

- Rapid congestion recovery (RCR): a better and quicker recovery of the motorway system during the recovery phase, which can be observed in system efficiency indicators, such as improved discharge flow rate, less total travel time and quicker mainline speed recovery.

Figure 1-1 is a conceptual cumulative volume curve for demonstrating these three key notions.
According to the statement of the research problem, the main research question is posed as follows:

When traffic demands taper off after the onset of motorway mainline congestion, how can RM be used to achieve rapid congestion recovery for better system efficiency during the recovery phase?

Under this main research question, the following more specific research questions have to be addressed:

1. What is the system benefit of RCR?
2. Can RM assist or accelerate motorway system recovery?
3. How can we identify the recovery phase after motorway mainline congestion?
4. How can we design strategies that use RM to achieve rapid congestion recovery?

1.3 AIMS

The aims of the research project are to investigate the feasibility of RCR using RM, and to develop strategies of RCR using RM. These aims are achieved through the following:

1. Review existing RM strategies in the literature to understand the current state of their art and practice.
2. Establish a coordinated RM system representing the current state of practice that is used as the bench mark for evaluating recovery strategy. This is achieved by developing a local queue management component and a coordination algorithm between on-ramps.
3. Explore the impact of the restrictive RM operation at the recovery phase.
4. Develop a RCR strategy using RM.
5. Evaluate the developed strategy and compare it with the former established bench mark in a micro-simulation environment.
6. Confirm the effectiveness of the proposed strategy in a modified cell transmission model (M-CTM) at a macroscopic level.
According to the literature review, the current state of practice for RM is a coordinated RM system. Therefore, this research requires a coordinated RM system representing the current practice as the benchmark for the proposed recovery strategy. The majority of the existing coordinated RM algorithms can be categorized into two approaches: model-based optimisation or rule-based heuristic. Almost all the coordinated RM systems in the fields are using rule-based heuristic approach, as the model-based optimisation approach requires too many real-time inputs which are not yet available in the field. In rule-based heuristic RM algorithms, some, including the Zone algorithm (Stephanedes, 1994), the Helper ramp algorithm (Lipp, Corcoran, & Hickman, 1991), SWARM (system wide adaptive ramp metering) (Paesani, Kerr, Perovich, & Khosravi, 1997) and the Bottleneck algorithm (L. Jacobson, Henry, & Mehyar, 1989), were developed in the 1980s and 1990s, so they are old and would not be appropriate representatives of the latest state of practice. Since 2000, two well-known field implementations of coordinated RM reported in the literature are HERO (heuristic ramp metering coordination) (Papamichail, et al., 2010) and SZM (stratified zone metering) (Geroliminis, Srivastava, & Michalopoulos, 2011). For HERO, a commercial system, there is not enough detail from the literature to fully model it. For SZM, the logic is much more complicated, and not enough detail can be found in the literature. Taking all these factors into consideration led to the choice to build a rule-based coordinated RM system targeting the state of current practice, and to use it as the benchmark.

1.4 CONTRIBUTIONS

This research project makes five major contributions:

1. Defined the recovery phase and analysed the feasibility of RCR using RM.

2. Developed a strategy for RCR using RM, called the zone-based RM strategy for RCR (ZRM-RCR); micro-simulation evaluation results for both the peak congestion scenario and the incident scenario suggested its effectiveness. Furthermore, an analysis of the macroscopic fundamental diagram (MFD) confirmed the increased merging capacity by ZRM-RCR at the recovery phase and the benefits of ZRM-RCR.

3. Developed a hybrid modelling framework for establishing the numerical relationship between the merging capacity and the metering rate for
metered ramps, and modified merge cells in the CTM by adopting variable merging capacity. This modified CTM (M-CTM) was then used for evaluating ZRM-RCR and the results further confirmed the effectiveness of ZRM-RCR.

4. Developed a multi-layer coordinated RM algorithm incorporating a PID feedback slave controller.

5. Developed a real-time on-ramp queue estimation algorithm, with high robustness, which is based on the Kalman filter theory.

1.5 THESIS OUTLINE

This thesis contains eight chapters presenting the development of a zone-based recovery strategy using RM. The content for each chapter is briefly outlined below.

Chapter 2 presents a comprehensive review of RM algorithms in the literature. These RM algorithms are analysed from a control system point of view. In particular, this chapter puts a special focus on issues related to field implementation of current practice. A brief review of Internet congestion control demonstrates its parallel: a major focus of Internet congestion control is recovery.

In Chapter 3, local queue management, a critical component for a field RM system, is presented. There are two main modules for local queue management: an on-ramp queue estimation algorithm provides real-time information; a queue management algorithm conducts adaptive metering control regarding the on-ramp queue length.

The next step for building a coordinated RM system, to develop the coordination between on-ramps, is presented in Chapter 4. The primary objective is to develop a coordinated RM algorithm which can overcome the limitations of local RM and so achieve network-wide benefits. The developed coordinated RM is used as the bench mark for the rest of the research project.

Discussions in Chapter 5 cover the critical concepts for recovery study. Firstly, the explicit definition of the recovery phase is given. Secondly, traffic dynamic features at the recovery phase are analysed, demonstrating potential benefits of rapid recovery at the recovery phase. Thirdly, an assumption is made that the restrictive metering control (RMC) can increase merge capacity, and the RMC is therefore set
as the basic RM operation for recovery phase. Fourthly, a micro-simulation exercise is conducted to support the assumption.

In Chapter 6, the development of ZRM-RCR and its evaluation in micro-simulation environment is presented. Results from two test-beds in the morning peak congestion scenario demonstrate the effectiveness of the proposed strategy. Additionally, the strategy is modified for an incident scenario, with the evaluation results also illustrating its effectiveness.

In order to further confirm the effectiveness of the proposed strategy, simulation results using an M-CTM are presented in Chapter 7. The M-CTM uses merge cells of variable merging capacity that is related to metering rate.

The conclusions of this thesis are summarised in Chapter 8.
Chapter 2  Literature Review

2.1  INTRODUCTION

The aims of this chapter are first to give a substantial analysis of the advancement of existing RM algorithms in the literature, and then to review briefly the classical algorithms in Internet congestion control that put recovery as an important objective. In particular, this chapter treats the RM system as a general control system and summarises the evolutionary progresses in RM systems from three components to form a control loop in Section 2.2. Then, a classification of RM based on its advancement, evolution of RM, is presented in Section 2.3. Section 2.4 summarises the state of practice of RM systems and identifies issues regarding field implementation. In Section 2.5, a brief review on Internet congestion control is reported. The implications and conclusions from the literature are summarised in Section 2.6.

2.2  EVOLUTIONARY PROGRESS IN RAMP METERING SYSTEMS

RM as a general control system has not only the control algorithm module, but has also a module for real-time system information obtaining and processing, as well as a signal timing module. The control loop of a RM system with the three modules is demonstrated in Figure 2-1. The first module retrieves and processes real-time information from the motorway network. With this real-time system information, it is then possible to apply reactive control: the more accurate the information is, the more suitable control actions can be made. The control algorithm can then decide suitable rates for metered ramps. The actuators at the end of the control loop finally realises the control and directly affect the motorway networks.

2.2.1  Efforts for Accurate Information in Real Time

The importance of real-time system information for a control system is self-evident. However, the first field implementation of RM signal ran pre-determined metering rates due to the unavailability of real-time traffic flow detection. Although pre-determined systems can improve merge condition by breaking the platoons, they cannot achieve high efficiency for motorway networks. The traffic detectors, such as the loop detector, are then developed to provide point data in real time that can partly
represent traffic flow information (Klein, Mills, & Gibson, 2006), giving the RM systems the complete control loop in Figure 2-1. The utilisation of traffic detector, identified as the first evolutionary progress in RM systems, advances RM from a pre-determined system to a reactive control system.

After the introduction of traffic detectors in RM systems, the efforts for the system information module focus on how to process detector measurements for accurate information. One example is for on-ramp queue estimation. As widely acknowledged, the most significant disadvantages of RM are the adverse impacts on ramp traffic and upstream surface streets. The way in which RM algorithms operate, restricting the entry of ramp traffic to the motorway mainline, creates on-ramp queues. Therefore, queue information is critical for a field RM system. Several studies (Liu, Wu, & Michalopoulos, 2007; M. Papageorgiou & Vigos, 2008; Vigos, Papageorgiou, & Wang, 2008) have proposed estimation algorithms based on detector measures and even on ramp signal timing information. Another recent example suggests using estimated density by probe vehicles instead of traditional point detector data (Kattan & Saidi, 2011). The recent rapid development and increased use of global positioning systems embedded in vehicles provides a possible way to obtain density in real time with reasonable costs. This would give opportunities for designing new RM algorithms.

The rapid development of vehicular communication cannot be ignored when talking about obtaining real-time information. With vehicular communication,
detailed individual vehicle information, such as position, speed, and destination, might change the way the ramp signals are controlled. For example, Park (2008) proposed a prototype algorithm, based on individual vehicle position and speed, which firstly assesses whether there are enough gaps for merging and how many vehicles can merge accurately, and then operates the green time of the ramp signal.

2.2.2 Efforts for Sophisticated Control Algorithms

The core of a control system is the algorithm which evaluates the system states and constrains in order to make control decisions. Particularly for RM algorithms, one clear evolution of RM algorithms can be seen in the literature is the influential scope when designing algorithms - at local or at network level. Local algorithms consider only local conditions and then decide metering rate. However, traffic conditions are not evenly distributed along the whole motorway network. Heavy ramp flows and lane reductions are the main causes of recurrent congestion. Traffic queues developed in such areas propagate upstream, activating additional bottlenecks. Given that a localised RM is operated, only the ramps located close to the bottleneck would take action to restrict the mainline access. Meanwhile, upstream ramps would not be efficiently used because they would not detect congestion with local information. Accordingly, one disadvantage of using only local information is the inefficient utilisation of ramp storage space. In addition, operating local RM independently leads to unevenly distributed ramp delays, which raises the so-called “equity” issue. The road users from the ramps that experience excessive ramp delays would strongly resist RM. Therefore, it is important to consider the motorway network as a whole system, determining the metering rates based on system-wide information. This made coordinated RM studies to be a hot topic in the 1990s (Lipp, et al., 1991; Paesani, et al., 1997; Stephanedes, 1994; Taylor, Meldrum, & Jacobson, 1998; Yoshino, Sasaki, & Hasegawa, 1995). Therefore, it is the second evolutionary progress – from local RM to coordinated RM.

Two objectives for RM systems conflict, especially during peak hours. The aim to maintain a free flow condition on the motorway mainline would naturally create long queues at on-ramps during peak hours. This would require the other objective, to shorten queue length so as to reduce the risk of queue spillover. In reality, even with ramp coordination these two conflicting objectives are impossible to be well balanced due to the excessive high peak demands. As a result, congestion would
sooner or later occur at motorways. Motorway congestion recovery, a valuable objective of RM, has no systematic research effort demonstrated in the literature. This is clearly a research gap for RM studies.

2.2.3 Efforts for Effective Actuators

The only available actuator in RM systems is the ramp signal. In other words, the only thing that RM systems can actually do is to change ramp signal timing so as to adjust merge traffic conditions. Advanced motorway management, therefore, requires actuators for motorway networks. Thanks to the advancement of intelligent transportation systems (ITS), many other control and management tools are now or will be available for employment in field operations: for example, dynamic route guidance systems (DRGS), variable speed limits (VSL), dynamic lane management systems (DLMS) and vehicular communication systems (VCS). As a result, RM itself has received increased emphasis in recent years under the umbrella of ITS. From a control system point of view, the cooperation of multiple actuators has the potential to maximise the efficiency and the capacity of existing motorway networks. Consequently, many recent studies start to model and test integrated ITS control systems (Carlson, Papamichail, Papageorgiou, & Messmer, 2009; A. Ghods & Rahimi-Kian, 2008; Hegyi, De Schutter, & Heelendoorn, 2003; Hou, Xu, & Zhong, 2007; Karimi, Hegyi, De Schutter, Hellendoorn, & Middelham, 2004; Lu, Varaiya, Horowitz, Su, & Shladover, 2011; Ni, Juan, & Zhu, 2008). Therefore, it is the third evolutionary progress in RM systems.

Other efforts have been made to improve ramp signal operation. The translation of the metering rate into traffic light settings, called metering operation policy, impacts directly on traffic flow, especially for merging. Papageorgiou and Papamichail (Markos Papageorgiou & Papamichail, 2008) summarised and compared the existing three types of metering operation policies:

- N-car-per-green, in which the green phase is fixed to allow N vehicles to enter per lane per cycle and the cycle time is calculated. One-car-per-green is the most popular metering rate because of its good merge performance, but it has a small maximum rate.
- Full traffic cycle, in which the cycle time is fixed and the green phase is calculated. It has a relatively high maximum rate but a large vehicle platoon.

- Discrete release rates, in which a set containing N fixed traffic light settings is provided. Therefore, only N fixed metering rates are available and the one that is closest to the metering rate ordered by the RM algorithm will be chosen.

2.3 EVOLUTION OF RAMP METERING

In this section, a classification based on the evolution of the RM systems is presented. Four ‘generations’ of RM algorithms are proposed (see Figure 2-2), ranging from early implementations to recent trends.

![Figure 2-2 Evolution of RM](image)

2.3.1 Pre-determined RM Systems

In a pre-determined RM system, metering rates are pre-determined according to clock time. Usually, the fixed plans are derived off-line based on constant historical demand data, without the use of real-time measurements. Due to their simple logic and the convenience of their implementation, these strategies were well-studied and implemented in the field in the 1960s and 1970s, confirming the first real use of RM (Pinnell, et al., 1967; Wattleworth & Berry, 1965). These pre-determined RM systems are classified as the first generation in the RM system evolution.
The advantage of this RM type is its easy implementation. Considering the capability of hardware and software in the 1960s and 1970s, this was the most cost-effective solution. However, the main drawback is also obvious – the solutions are based on historical data rather than on real-time data. This is a crude simplification for the following reasons: firstly, demands are not constant and vary significantly over different days; secondly, incidents and further disturbances may affect traffic conditions in unpredictable ways. Hence, pre-determined RM systems may lead to either overload of the mainstream flow or underutilization of the motorway. Nowadays, these early systems almost not applied in the field; they are used as a backup mode which would be activated on rare occasions, such as where there is a lack of real-time measurements or a deficiency of detectors.

2.3.2 Local Responsive RM Systems

Unlike in pre-determined RM systems, a metering rate for an on-ramp in a local responsive RM system is determined based on its local traffic flow measures, such as flow, occupancy or speed, and occasionally queue overflow on the metered ramp, from adjacent upstream or downstream detectors. Examples include the demand-capacity (DC) and the occupancy (OCC) strategies (Masher et al., 1976), RWS algorithm (Middelham & Taale, 2006), the local fuzzy logic algorithm (Taylor, et al., 1998), ALINEA (Papageorgiou, Hadj-Salem, & Blosseville, 1991; Papageorgiou, Hadj-Salem, & Middelham, 1997), the neural network (NN) algorithm (H. Zhang & Ritchie, 1997) and ANCONA (B.S. Kerner, 2005). The responsive approach is more advanced than the pre-determined approach because of its responsive nature of control philosophy which reacts to and adjusts the metering rate according to the varying traffic situation in real time. This type of RM system can be considered as a complete control system. Due to the development of motorway infrastructure and information techniques, local responsive RM systems have been widely applied in the field since the 1980s. Therefore, local responsive systems are classified as the second generation.

For this RM type, DC, OCC and the local fuzzy logic or RWS algorithms are feed-forward schemes, while ALINEA, local NN and ANCONA algorithms are feedback schemes. Whilst the DC algorithm is very popular in North America, ALINEA is the most successful local RM algorithm due to its simplicity and
classical feedback nature (M Papageorgiou, et al., 1997). Both are considered to be examples of local responsive RM.

**Demand-Capacity (DC) Algorithm**

The DC algorithm calculates the metering rate \( r(k) \) at the kth interval by a pre-specified capacity and measured incoming flow rate at the (k-1)th interval:

\[
\begin{align*}
    r(k) &= \begin{cases} 
    q_{cap} - q_{in}(k-1), & \text{if } o_{out}(k) \leq o_{cr} \\
    r_{min}, & \text{else}
    \end{cases} 
\end{align*}
\]

(2-1)

where \( q_{cap} \) is the motorway capacity downstream of the ramp;

\( q_{in} \) is the motorway flow measurement upstream of the ramp;

\( o_{out} \) is the motorway occupancy measurement downstream of the ramp;

\( o_{cr} \) is the critical occupancy, where motorway flow becomes maximum;

\( r_{min} \) is a predefined minimum metering rate.

The DC algorithm is a feed-forward control which adjusts control input \( (r(k)) \) according to measured disturbance \( (q_{in}) \) rather than control outcome \( (o_{out}) \), so it is blind to the control outcome, and is generally known to be quite sensitive to the non-measured disturbances.

OCC and RWS algorithms are based on the same philosophy as the DC algorithm. However, OCC uses occupancy measurement to estimate \( q_{in} \), while RWS employs speed measurement to switch RM on / off and to decide whether the traffic condition is over the critical point (the point in the fundamental diagram where motorway flow achieves maximum) or not.

**ALINEA**

Unlike feed-forward schemes, ALINEA’s closed-loop feedback control structure makes it the most successful local RM. Feedback, one of the most classical control philosophies, has many advantages. Firstly, ALINEA is less sensitive with respect to disturbances due to the robustness of the feedback control. Secondly, ALINEA reacts smoothly, and therefore may prevent congestion by stabilising the traffic flow at a high throughput level. Thirdly, ALINEA is very simple and computationally efficient. In ALINEA, the metering rate at the kth interval is calculated as follows:
\[ r(k) = r(k - 1) + K_r[\hat{\delta} - o_{out}(k)] \]  

(2-2)

where \( K_r > 0 \) is a regulator parameter (integral gain);

\( \hat{\delta} \) is the desired occupancy of the downstream detector;

\( o_{out} \) is the downstream occupancy measurement.

The control objective of ALINEA is to maintain the downstream occupancy around the desired occupancy so that the downstream flow can keep close to its capacity. Occupancy is used as the feedback variable because it is relatively stable in weather and stochastic conditions. From the control point of view, ALINEA is a discrete-time I-type (Integral-type) controller. The integral term responses to time-Integration of the control error \( (\hat{\delta} - o_{out}(k)) \), representing the past situation. Therefore, it has a zero steady-state error but a relatively slow control response. In the case of a merge section bottleneck, most vehicles released from the on-ramp will reach the downstream detector location within an update interval \( T \) so ALINEA can react properly. However, the performance of ALINEA will be reduced in case of the distant downstream bottlenecks because of the longer time delay between the control input (on-ramp flow) and its impact at the downstream detector location. Therefore ALINEA is extended to a PI-type (Proportional-Integral-type, where proportional term is like to response to the current situation) controller in which a proportional term is added to speed the control response (Wang & Papageorgiou, 2006). The PI-type ALINEA is shown as follows:

\[ r(k) = r(k - 1) + K_r[\hat{\delta} - o_{out}(k)] - K_p[o_{out}(k) - o_{out}(k - 1)] \]  

(2-3)

where \( K_p > 0 \) is the proportional gain and the rest of the parameters have the same values as in Equation 2-2.

A Local NN algorithm is similar to ALINEA, but uses an NN to mimic an I-type controller or PI-type controller. In ANCONA, congestions are allowed to happen (which makes it different to other RM algorithms which seek to avoid congestion); RM is then activated to relieve the congestion (this switching rule is called bang-bang control). ANCONA is proposed based on three-phase traffic theory (Boris S Kerner, 2004).
2.3.3 Coordinated RM Systems

As discussed in Section 2.2.2, the most significant drawback of local RM is the lack of system-wide information. Consequently, the natural solution is to consider the whole network as a system when applying RM: that is, coordinated RM. Coordinated RM systems make use of measurements from an entire motorway network to control all metered on-ramps for the optimal solution of the whole motorway network. Compared with local responsive RM systems, this RM type has the potential to make full use of all ramp storage spaces, and to take user equity into consideration. Coordinated RM is thus an improvement over local RM systems. Development of computer and information techniques allows the real-time monitoring and the control of the traffic signals of a large network in a control centre. Therefore, coordinated RM systems are classified as the third generation in the RM evolution.

Coordinated RM studies have been undertaken extensively over the last three decades. Many coordinated RM algorithms have been proposed in the literature, and a few have installed in the field (mainly in US and Europe). This study further subdivides these algorithms into two categories according to the different ways of determining the coordination between ramps: rule-based heuristic algorithms and model-based optimisation algorithms.

Rule-based Heuristic Algorithms

The coordination in rule-based heuristic algorithms is determined by predefined heuristic rules. There are many examples and most of them have already been implemented in the field, including the ZONE algorithm (Stephanedes, 1994), the Helper ramp algorithm (Lipp, et al., 1991), the Linked-ramp algorithm (Banks, 1993), the Bottleneck algorithm (L. Jacobson, et al., 1989), the system wide adaptive ramp metering (SWARM) (Paesani, et al., 1997), ACCEZZ (adaptive and coordinated control of entrance ramps with fuzzy logic) algorithm (Bogenberger, Vukanovic, & Keller, 2002), the Sperry algorithm (Virginia Department of Transportation, 1998), SZC (Geroliminis, et al., 2011) and the HERO algorithm (Papamichail, et al., 2010). Table 2-1 lists these algorithms and their implementation status. Rules here are usually based on simple principles or are extracted from historical analysis and expert experiences. For example, the demand-capacity principle, which has been used in many coordinated RM algorithms, requires that
total traffic demands in an area (containing several on-ramps) for a bottleneck (including mainstream flow and ramp flows) should be no more than the bottleneck capacity. Another example is the Fuzzy logic technique based algorithm ACCEZZ. Based on historical analysis and expert experience, it would generate several traffic condition patterns incorporating system-wide information for each on-ramp and would pre-define a metering rate for each pattern.

The advantages of this RM type are as follows. Firstly, from an engineering perspective, rule-based heuristic algorithms are easy to implement, so most rule-based RM algorithms have been operating in the field. Secondly, no motorway model is needed, so computations are efficient. Thirdly, the expertise of traffic engineers can be incorporated in the rule definitions.

The disadvantages include the unclear theoretical foundation and the costs of the parameter fine-tuning. More precisely, the effectiveness of most heuristic algorithms cannot be proved theoretically, so solutions from these algorithms may not be optimal. Additionally, simple rules are always of a feed-forward nature, which cannot accurately describe complicated traffic conditions, especially for critical conditions. For instance, computing the metering rate based on demand-capacity principle requires capacity as a given input, and it is widely acknowledged that capacity is stochastic. Therefore, using a fixed capacity always causes a mismatch with the reality: this makes effectiveness rely heavily on the selection of parameters, which increases the costs of tuning these parameters.

**Model-based Optimisation Algorithms**

In model-based optimisation algorithms, the metering rates over an optimization time-horizon are calculated by a numerical optimization algorithm based on the macroscopic model. From the control point of view, an RM problem can be seen as a system optimal problem. The motorway network model can be described in a state-space form:

\[ x(k + 1) = f [x(k), u(k), d(k)] \] (2-4)

where \( x(k) \) is the network state vector, including occupancy, mean speed and queue length;
$u(k)$ is the control input vector, including the metering rates of metered on-ramp;

$d(k)$ is the external disturbance vector which is usually the traffic demand.

Table 2-1 List of rule-based heuristic algorithms

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Implementation status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zone algorithm</td>
<td>The motorway network is divided into zones which end at a bottleneck. The algorithm aims at balancing the entering and exiting traffic volumes of each zone.</td>
<td>Minneapolis / St. Paul, Minnesota</td>
</tr>
<tr>
<td>Helper ramp algorithm</td>
<td>For the coordination aspect, once a ramp is classified as critical, the Helper algorithm immediately begins to override upstream ramp control to reduce upstream metering rates.</td>
<td>Denver, Colorado</td>
</tr>
<tr>
<td>Linked-ramp algorithm</td>
<td>The coordination aspect of this system rests on a heuristic logic, similar to that of the Helper algorithm.</td>
<td>San Diego, California</td>
</tr>
<tr>
<td>Bottleneck algorithm</td>
<td>At a network-wide level, the formation of congestion at various bottleneck locations is identified and a decision is made with respect to the required volume reduction.</td>
<td>Seattle, Washington</td>
</tr>
<tr>
<td>SWARM</td>
<td>A linear regression and a Kalman filter are applied to detector data for the forecast of future traffic demands.</td>
<td>Orange County, California</td>
</tr>
<tr>
<td>ACCEZZ</td>
<td>The rule base, defined as the set of rules in the fuzzy logic algorithm, incorporates human expertise.</td>
<td>Germany</td>
</tr>
<tr>
<td>Sperry algorithm</td>
<td>The strategy operates at two distinct modes, the restrictive one and the non-restrictive one, with respect to a predefined threshold.</td>
<td>Arlington, Virginia</td>
</tr>
<tr>
<td>HERO</td>
<td>When the queue of an on-ramp becomes larger than a predetermined threshold, the burden of decreasing this queue is assigned to upstream on-ramps.</td>
<td>Melbourne, Victoria</td>
</tr>
</tbody>
</table>

The cost criterion is usually an efficiency indicator, like travel time spent (TTS) or total waited time (TWT). The real world constraints, like queue length and minimum / maximum metering rate, are also formulated as the penalty terms. The objective function is finally obtained by combining the chosen cost criterion and the penalty terms. The RM problem is then transformed to an optimisation problem which can be solved by optimal control techniques, such as linear quadratic programming or nonlinear programming.
Examples of this RM type include the Linear Programming algorithm (Yoshino, et al., 1995), the optimal advanced system for integrated strategies (OASIS) (Kotsialos, Papageorgiou, Mangeas, & Haj-Salem, 2002) and the advanced motorway optimal control (AMOC) (Kotsialos & Papageorgiou, 2004). The details of these algorithms are listed in Table 2-2, including name, description and implementation status.

Table 2-2 List of model-based system optimization algorithms

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Implementation status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Programming</td>
<td>This algorithm is based on linear programming formulation and maximizing the</td>
<td>No</td>
</tr>
<tr>
<td>Programming algorithm</td>
<td>weighted sum of the ramps. Weights are selected to reflect the relative</td>
<td></td>
</tr>
<tr>
<td></td>
<td>importance of the ramps.</td>
<td></td>
</tr>
<tr>
<td>AMOC</td>
<td>AMOC employs the METANET model to formulate coordinated RM as a discrete-</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>time non-linear optimal control problem and uses a feasible-direction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>algorithm to obtain a numerical solution over a given time horizon.</td>
<td></td>
</tr>
<tr>
<td>OASIS</td>
<td>OASIS is similar to AMOC but employs the METACOR model.</td>
<td>No</td>
</tr>
</tbody>
</table>

In order to make this type of algorithm more proactive and applicable, some further literature (Bellemans, De Schutter, & De Moor, 2006; De Schutter, Hellendoorn, Hegyi, van den Berg, & Zegeye, 2010; Gomes & Horowitz, 2006; Hegyi, et al., 2003) suggests embedding the algorithm in a model predictive control (MPC) structure. MPC is an advanced method of process control that has been used in process industries such as chemical plants and oil refineries since the 1980s (Camacho & Bordons, 1999). In MPC, an estimation / prediction layer is added which uses real-time measurements to reduce mismatches and to improve model accuracy. In every update interval (h), the optimal solution is calculated over an optimization horizon (H), but when applying the optimal solution, only the results in the rolling horizon (h<H) are used. The same process is repeated in the next update interval. Within the MPC structure, the model-based optimal algorithms are extended to a hierarchical control system which consists of three layers (see Figure 2-3): the estimation / prediction layer, the optimization layer and the direct control layer. The estimation / prediction layer receives historical data, incident information and real-time measurements from detectors as input, and then produces the current state
estimation and the prediction of future disturbances. In the optimization layer, the model-based optimal algorithm is fed with information from the estimation / prediction layer to obtain optimal results. The direct control layer consists of independent regulators (one for each metered ramp) that use the optimal results as the set values for their operation.

Figure 2-3 Model predictive control structure (Papamichail, Kotsialos, Margonis, & Papageorgiou, 2008)

There are two major advantages of using model-based optimization algorithms. Firstly, compared with rule-based heuristic algorithms, model-based optimization algorithms have a good theoretical foundation. Theoretically, the solution is the global optimal solution if the problem can be solved. Secondly, it is a proactive solution because all the disturbances are assumed to be known over the optimization time-horizon, guaranteeing congestion avoidance at a maximal level theoretically.

However, the disadvantage of this type is the difficulty of applying it in the field in real time, even in a MPC structure. It is almost impossible in the field to feed the optimisation model with so many real-time system measurements required. Even though all the inputs can be estimated, the accuracy of the traffic flow model is still questionable, especially for the transitions between free flow and congestion, making it difficult to apply those algorithms in practice.
2.3.4 RM integrated in ITS

As noted in Section 2.2.3, the current trend of considering RM under the umbrella of ITS is because RM alone is incapable of providing consistent high level mobility to motorway users, especially during peak hours. There are three limitations to RM. Firstly, RM achieves most of its benefits when demand is higher than capacity in limited areas during limited periods (Sawyer, 2002). Secondly, RM has limited direct impact on mainline traffic flow, especially for the long mainline section. Thirdly, all the existing RM algorithms are using aggregated data from point detectors, so RM cannot guarantee every individual vehicle enough merge gaps. However, with the rapid development of ITS, many other control tools, such as DRGS, VSL or DLMS and VCS, are now or will be available for employment in field operations. These tools have the potential to shift peak time demand, or directly impact mainline traffic flows, and to provide individual vehicular information in real time. Therefore, its integration with other management tools in ITS is a new opportunity which will enable RM to make significant improvements in traffic congestion. Some recent research (Carlson, et al., 2009; A. Ghods & Rahimi-Kian, 2008; Hegyi, et al., 2003; Hou, et al., 2007; Karimi, et al., 2004; Kattan & Saidi, 2011; Lu, et al., 2011) has already illustrated through preliminary simulation results that the integration of RM and other ITS tools is promising. Therefore, RM integrated in ITS is classified as the fourth generation of RM, representing the future RM in ITS. There are three levels at which RM can be integrated in ITS: macro, meso and micro.

Integration at Macro Level

The fundamental cause of congestion is that traffic demand exceeds infrastructure capacity. Travel demand management, which is defined as providing travellers with effective choices to improve travel reliability, is at the fundamental level in ITS. Travel demand management can influence not only motorway traffic but also arterial traffic, and public transportation systems. Therefore, integrating RM with travel demand management tools at a macro level is important.

One typical tool of travel demand management is DRGS, which employs public communication means (such as radio and web services) to inform drivers about current or expected travel times and queue lengths so that they may reconsider their route. DRGS can offer both pre-trip recommendations and en route advice, so it
can influence the demand on the whole transport system. With the help of DRGS, the demand during peak hours can be distributed more equally. As a result, less field constraints for RM system would be activated, so RM system can achieve its highest benefits.

**Integration at Meso Level**

A meso level integration means that RM is only cooperating with other motorway control tools, such as VSL and DLMS. Compared with integration at a macro level, integration at a meso level includes only control tools for motorways, and affects only motorway traffic.

As an access control on motorways, RM has direct impact only on the traffic flow in the area adjacent to the on-ramp. Consequently, RM alone is insufficient for establishing optimal traffic conditions (maximum efficiency) for the whole motorway. For example, when the congestion is caused by large ramp flows or merge-bottlenecks, RM is effective in alleviating congestion. However, RM is not as effective for congestion caused by large mainline flows. In this situation, VSL is useful because it can maintain high density and stable flows in the mainline (Bertini, Boice, & Bogenberger, 2006; Highway agency, 2004), and so the integration of RM and VSL is able to achieve higher efficiency for the whole network. On the other hand, these tools can provide better near ramp traffic conditions for RM. For instance, DLMS can direct more vehicles to change to the middle lanes so that more space in the kerbside lane will be available for vehicles from the on-ramp to merge.

**Integration at Micro Level**

At the micro level, metering operation policy is investigated. In existing field RM algorithms, a fixed metering policy translates the determined metering rates into traffic light settings. This may lead to the following results: some merging vehicles may be held due to a lack of sufficient gaps, thus resulting in irregular merging and potential disruptions to mainline traffic; and others may have too much room, resulting in unused capacity. Therefore, the ideal metering policy is a dynamic one in which ramp signals will discharge a variable number of vehicles according to the available gaps in the merge area.

Aggregated data from point detectors does not achieve an ideal metering policy. However, this may be possible in the future with the help of vehicle
infrastructure integration (VII) systems. On one hand, VII can provide more detailed data for each individual vehicle including the current / previous location of the vehicle, instantaneous states such as velocity and acceleration / deceleration, intention for turning movement, and vehicle spacing (Park, 2008). With this information, the ramp signal can accurately assess whether there are enough gaps for merging and how many vehicles can merge. On the other hand, VII can directly inform drivers by vehicle to infrastructure (V2I) communication whether they can pass through current or incoming green phases or not, so the dynamic metering policy can be achieved.

2.4 RM SYSTEMS IN PRACTICE

2.4.1 State of Practice

RM is probably the first motorway control tool that has been implemented in the field long before ITS. Since the first metered ramp in Chicago on the Eisenhower Expressway in 1963 (Pinnell, et al., 1967; Wattleworth & Berry, 1965), over 50 metropolitan areas have deployed RM systems worldwide. In North America, at least 33 cities are running RM systems and the estimated number of metered ramps is approximately 2600 (Piotrowicz & Robinson, 1995; US Department of Transportation, 2006). Europe is another area where RM systems have been applied widely to improve motorway traffic. There is no exact figure indicating the number of current systems and metered ramps, but the literature (Middelham & Taale, 2006; Muhurdarevic, et al., 2006) indicates that RM systems are widely applied in France, Germany, England and the Netherlands. In Australia, capital cities are interested in RM; Melbourne (Papamichail, et al., 2010) has already applied RM, and Brisbane has already commenced field tests.

All these field RM systems are nowadays running traffic responsive control systems. Local responsive RM is the basic module. Most of these have deployed one or more coordination algorithms on top of the local RM; the rest are considering upgrading to coordinated systems. In particular, rule-based coordination is dominating field installations. In summary, the field systems are now at the third generation of the RM evolution.
2.4.2 Benefits of RM Implementation

Many reports (Cambridge Systematics Inc., 2001; Cassidy & Rudjanakanoknad, 2005; M Papageorgiou & Kotsialos, 2002; Piotrowicz & Robinson, 1995) have been written which document the success and list the benefits of implementing RM. In general, RM systems are able to improve system efficiency, increase safety and benefit the environment. Therefore, three types of performance indicators are usually used to evaluate RM applications:

- System efficiency indicator (SEI) is the most important and widely-used measure to evaluate RM. Three SEIs are commonly used. Total time spent (TTS) / vehicle-hours travelled (VHT) is a measure of overall system performance. All vehicles, including those having finished their journey and those currently on the motorway, are considered in this measure. Throughput, another measure of overall system performance for the whole network, measures the total number of vehicles served by a bottleneck or the whole network. Average mainline travel speed (AMTS) is a measure of traffic conditions on the motorway mainline.

- Safety indicator is used to evaluate safety in merging area at on-ramps with RM. Incident rate (IR) at merging areas is one commonly used safety indicator. In field tests, IR can be calculated, but a large amount of data is required for this and it may take a long time to collect a sufficient sample size. Furthermore, in simulation situations, it is difficult to calculate IR because there is no random incident in the simulation.

- Environment indicator usually estimates the fuel consumption and emissions of the total vehicles. With improved traffic flow on motorways, vehicles’ fuel consumption and emissions will be reduced, particularly with the reduction of stop-and-go conditions in the mainline. On the other hand, more congestion at on-ramps creates more fuel consumption and higher vehicle emissions.

The field evaluation results prove that RM is an effective control tool for motorway networks with the following benefits:

- Improve the efficiency of motorway capacity usage;

- Improve the safety and reliability of the motorway system;
- Decrease congestion and shockwave propagations caused by merging platoons;
- Reduce vehicle emissions on motorway mainlines.

2.4.3 Issues Raised by Field Implementation

Apart from the benefits, there are several issues raised by the field implementation of RM. This sub-section summarises three issues as the feedback of field implementation. These issues discuss the side-effects and the costs of RM, which can help motorway operation agencies select proper RM systems and evaluate the cost.

**User Equity**

User equity is an issue that is often at the heart of public opposition to the implementation of RM. Simply stated, RM provides the most benefits (travel time saving) to long-distance travellers who are passing through the metered on-ramps. Consequently, close-in or city residents, especially in the metered areas, feel an inequity about this, as they are bearing the expense of the benefits to the long-trip travellers by having limited access to and increased delay at the metered on-ramps.

It is difficult to determine an indicator for user equity because this concept is very difficult to define. Whilst a large amount of research has been devoted to the measurement from an economic equity perspective (like Gini coefficient), Zhang and Levinson (Levinson & Zhang, 2005) may be the first to introduce income inequity measures to RM. According to their research, there are two types of user inequity indicators: temporal and spatial. Temporal inequity is the difference in travel time, delay or speed among drivers who travel on the same route but arrive at the on-ramp at different times, while spatial inequity is the difference in travel time, delay or speed among drivers who arrive at different on-ramps at the same time. Typically, the base scenario, which represents perfect user equity, is the traffic flow situation without any RM.

**On-ramp Queue Management**

On-ramps are usually the buffers between motorways and arterials. On arterials, platoons are encouraged to improve the quality of flow. On motorways, platoons can break the traffic flow at a merge. On-ramps are the links between these...
two different systems. RM acts as transitioning elements to split-up platoons and prepare vehicles to merge with motorway flow conditions. As a result, RM operations might affect arterials significantly once the buffer has been consumed.

It is acknowledged that the most severe disadvantage of RM is the ramp queue spillover onto arterial roads. In the field, queue spillover is usually forbidden by disabling RM to allow more vehicles entering motorways. Consequently, a free flow condition cannot be maintained on motorway mainline, and the benefits of RM are reduced. There are two considerations for queue management. One is at the very beginning of installing RM the selected ramps should have some storage spaces for a better buffering. The other is to have a queue management system that makes a sophisticated and fully functional RM system.

**System Maintenance**

Metered ramps would diverge freeway trips to adjacent surface streets as some drivers may choose other routes to avoid ramp queuing. Many evaluations of existing RM systems suggest that adjustments in traffic patterns after RM is implemented take a few months and are in many forms. Besides, the motorway mainline traffic flow patterns will also be changed after the implementation of RM. For example, increased traffic volume through a previous bottleneck could create new bottlenecks in its downstream. All these changes of traffic flow patterns must require corresponding changes of the RM system settings. These changes may include recalibration of algorithm parameters, improvement of strategy and installation of new infrastructure (such as new variable message boards).

### 2.4.4 Lessons Learned from Practice

As discussed above, a real RM project includes many aspects, far more than just the algorithm development. One objective of the literature review is to assess the RM related studies in the literature from a control system point of view. The focus is still on academic studies rather than on engineering in industries. Therefore, this subsection summarises the lessons learned from practice that support RM algorithm development and improvement.

The key word, simplicity, can be elaborated into two aspects. Firstly, the concept of a successful algorithm should be easy to understand. Secondly, the relationships between components in the algorithm should be clear. For example,
ALINEA is simple but the most successful RM algorithm. What is the reason? The acceptance of engineers who are conducting the field RM implementation is critical for a successful RM application. On the one hand, the best way to convince agencies and engineers is to make them understand it. On the other hand, a RM system implementation is not a simple work. It requires many engineers to participate and takes years. As just noted in Section 2.4.3, this process would involve re-calibration of parameters and algorithm improvement. If it is difficult to understand, it would be difficult for engineers to calibrate and improve the algorithm.

2.5 INTERNET CONGESTION CONTROL

2.5.1 Why Internet Congestion Control

A search on the Internet for the keywords “traffic congestion control” presents results that mostly pertain to Internet congestion rather than to roadway network congestion. The speed of increasing demand for the Internet is much faster than its counterpart in motorway traffic. As is widely acknowledged, the structure of the Internet is extremely complicated, so the currently implemented congestion control strategies must be simple to apply and sophisticated enough to alleviate Internet congestion. Therefore, Internet congestion control theories might be applicable to roadway networks with appropriate modifications, offering new ways to envision roadway traffic management. This is the objective of this investigation of Internet congestion control.

This section firstly gives an overview and classification of Internet congestion control. Algorithms in TCP (transfer control protocol) congestion control are then summarized.

2.5.2 Overview

Congestion control mechanisms in today’s Internet already represent one of the largest deployed artificial feedback systems (Low, Paganini, & Doyle, 2002), due to the mechanism of acknowledgement (ACK) in sending data packet, in which the sender needs an ACK from the receiver to confirm the end of sending one data packet.
In Internet congestion control, almost all the implemented algorithms can be classified as end-to-end congestion control. The end-to-end principle is one of the central design principles of the Internet and is implemented in the design of the underlying methods and protocols in the Internet Protocol Suite. The principle states that, whenever possible, communication protocol operations should be defined to occur at the end-points of a communication system, or as close as possible to the resource being controlled. The wide usage of end-to-end congestion control mechanisms has been a critical factor in the robustness of the Internet (Floyd & Fall, 1999).

There are two primary components in congestion control: the source algorithm executed by host computers and edge devices, and the link algorithm executed by network devices.

The source algorithm is the most widely used Internet congestion control algorithm in which the source device, such as host computers, determines sending rate to the network. This is quite similar to the installation of ramp signal that targets the control of vehicles entering motorway networks. Algorithms in TCP congestion control, the major one in use today, are similar to RM and will be introduced in Section 2.5.3.

The link algorithm, actually the queue management strategy for the network device, such as routers and servers, is not similar to RM, so is not discussed.

2.5.3 TCP Congestion Control

TCP, a core protocol of the Internet Protocol Suite, is one of the two original components of the suite, complementing the Internet Protocol, and therefore the entire suite is commonly referred to as TCP/IP. TCP provides reliable, ordered delivery of a stream of bytes from a program on one computer to another program on a different computer. TCP is the protocol that major Internet applications rely on, applications such as the World Wide Web, e-mail, and file transfer (Postel, 1981).

The traditional end-to-end congestion control mechanisms of TCP, which employs an additive increase multiplicative decrease (AIMD) algorithm (V. Jacobson, 1988), have been a critical factor in the robustness of the Internet. Modern implementations of TCP contain four algorithms: slow-start, congestion avoidance,
fast retransmit and fast recovery. The first two algorithms, slow-start and congestion avoidance, are basic algorithms which are detailed revealed.

**Slow-start**

In Internet packet transmission, “self-clocking” is based on packet conservation and the ACK feedback mechanism, which means equilibrium will be reached and the packet transfer speed will be adjusted automatically by the receiving ACK. The problem is how to get this “self-clocking” at the beginning of one connection. To start the “clock”, the slow-start algorithm is developed to gradually increase the amount of data in transit (V. Jacobson, 1988). The algorithm is subtle, expressed using pseudocode as follows:

- Add a congestion window, "cwnd", to the pre-connection state.
- When starting or restarting after a loss, set cwnd to one packet.
- On each ack for new data, increase cwnd by one packet.
- When sending, send the minimum of the receiver’s advertised window and cwnd.

With this algorithm, a source can quickly reach its maximum bandwidth in a short time, which is actually “quick start”.

**Congestion Avoidance – AIMD Algorithm**

Congestion avoidance is also called the AIMD algorithm. The AIMD mechanism is primarily intended to address computer network congestion promptly by setting the entering network traffic rate to a level that is sufficiently low to quickly recover the network and avoid further deterioration of the network traffic flow. This concept is also important for roadway traffic networks because traffic queues can form quickly at bottlenecks but take much longer to discharge.

The logic of AIMD algorithm is very simple:

- On congestion (like, there is no ACK received when timeout is over):
  
  \[
  cwnd(i) = d \cdot cwnd(i - 1) \quad (d < 1) \quad (2-5)
  \]

- No congestion (like, receive three consecutive ACKs):
  
  \[
  cwnd(i) = cwnd(i - 1) + u \quad (u \ll cwnd_{\text{max}}) \quad (2-6)
  \]
where cwnd is congestion window size;

d is the multiplicative gain;

u is a constant increment.

From the logic, the algorithm actually reacts to congestion and tries to speed recovery by reducing the congestion window size multiplicatively.

The algorithm is expressed using pseudocode as follows:

- On any timeout, set cwnd to half the current window size (this is the multiplicative decrease).
- On each ack for new data, increase cwnd by 1/cwnd (this is the additive increase).
- When sending, send the minimum of the receiver’s advertised window and cwnd.

Note that this algorithm is only congestion avoidance, and it does not include the previously described slow-start. Since the packet loss that signals congestion will result in a re-start, it will almost certainly be necessary to slow-start in addition to the above. So the slow start algorithm and AIMD are always combined together. In TCP, this combination is as follows:

- if (cwnd < ssthresh):

  /* if we’re still doing slow−start */
  /* open window exponentially */
  cwnd += 1;

- else:

  /* otherwise do Congestion */
  /* Avoidance increment−by−1 */
  cwnd += 1/cwnd.
2.6 SUMMARY

This chapter highlights two viewpoints found when reviewing existing RM studies in the literature. One is the control system point of view. From this point of view, RM systems are analysed based on major advancements. Accordingly, a classification of RM systems is proposed, namely the evolution of RM. The other one is the practice point of view. The ultimate goal of RM study is to develop a RM system that can be applied in the field for better managed motorways. The practical viewpoint provides an investigation of real status and problems for RM field installation, and then research gaps with practical value can be identified.

The review of RM draws two research gaps from these two viewpoints. One is to consider RM in ITS with other motorway management tools. The other one is to investigate RM for rapid congestion recovery (RCR). The latter one is of current practical value, and thereby is selected as the research problem for this thesis.

Another conclusion drawn from this review is that the coordinated RM system is of current practice state. Therefore, this research will use a coordinated RM system as the bench mark for evaluating recovery strategies using RM.

A brief review of Internet congestion control identifies the similarity between RM and the source algorithms of TCP congestion control. Moreover, TCP congestion control reacts to congestion by the loss of ACK, and then activates actions for recovering the network. This suggests implications for motorway congestion recovery.
Chapter 3  Local Queue Management

3.1  INTRODUCTION

The purpose of this chapter is to present relevant works that have been carried out for developing a local ramp metering (RM) sub-system with queue management. As noted in Section 2.4.3, on-ramp queue management is a must-have component for field RM system. Therefore, this part of study is an important work for establishing the bench mark – a coordinated RM.

The task of local queue management is to reduce the on-ramp queue length and the risk of queue spillover. A commonly used on-ramp queue management strategy is the so-called “queue flush or override” method. This method is enabled by placing a detector close to the upstream end of the ramp. The measured detector occupancy exceeding a threshold indicates that the queue has reached the detector location, and the metering rate is increased to the maximum level to clear the traffic queue. Although this simple method can quickly relieve queue spillover, sudden increases and oscillations in the metering rate adversely affect the mainline traffic and diminish the main purpose of RM (Chaudhary, Tian, Messer, & Chu, 2004; Tian, Messer, & Balke, 2004).

A recent study (Spiliopoulou, Manolis, Papamichail, & Papageorgiou, 2010) proposed an on-ramp queue controller which could regulate on-ramp queues smoother with both real-time queue estimation and on-ramp demand prediction one time-step ahead. Although the controller is much more sophisticated than simple “queue flush”, the queue controller simply overrides the normal metering control to put the on-ramp queue length as a new control objective when activated. There is no mechanism to balance between the two conflicting objectives, so this arbitrary replacement of control objective could still lead to adverse impacts on mainline traffic, reducing the initial benefits of applying RM.

The above analysis reveals that accurate queue estimation in real time and a mechanism to balance RM and queue management are the keys to successful local queue management. Consequently, an on-ramp queue estimation algorithm and an on-ramp queue management scheme are the two focuses in this chapter.
3.2 FRAMEWORK

The purpose of this section is to build up the framework of the local RM sub-system with queue management (see Figure 3-1). There are two ways of generating the metering rate. One is to calculate it by a basic metering algorithm. This research takes ALINEA as the basic local RM algorithm. The other way is to generate it by queue management algorithm based on a real-time queue estimation. Then, a decision making module with the mechanism of balancing ALINEA and queue management objectives determines the final rate.

![Figure 3-1 Local RM sub-system](image)

3.3 QUEUE ESTIMATION

Sophisticated queue management relies on accurate queue information in real time, so the primary objective of this research is to develop a robust queue estimation algorithm for motorway on-ramps. The proposed algorithm is developed based on the Kalman filter framework. Fundamental traffic flow conservation is used to estimate the system state (i.e., queue sizes) based on the flow-in and flow-out measurements. This estimation inevitably produces noise due to the counting error. Therefore, the estimated results are updated with the Kalman Filter measurement, which uses loop detector time occupancies. This section also proposes a novel single point correction. The method resets the system state when a significant change in the mid-link time occupancy is observed for eliminating the accumulated counting error.

After a simple review of existing techniques, the development and the evaluation of the algorithm is presented.
3.3.1 Existing techniques for Ramp Queue Estimation

Given the motorway on-ramp as the applying object and the conventional loop detector as the information source, recent queue estimation studies can be categorised into two types. The first approach uses the flow conservation model with flow-in and flow-out counts. This method assumes the traffic flow conservation rule and estimates the number of vehicles in a given roadway segment by calculating the difference of flow-in and flow-out counts measured by two detectors placed at the entrance and exit of the link. Although this approach is easy to understand and implement, a critical drawback is the detector counting error that accumulates over time. Even with a reasonable level of error rate, the accumulative errors may render the estimation results useless. Liu et al. (Liu, et al., 2007) reported their study in Minnesota that 31% of metered ramps have biased loop detectors in counting number of vehicles. The conservation model could be applied to only 60% of ramps because of the counting noise of loop detectors.

A second approach uses advanced techniques to correct the estimation results of the first approach. Liu et al. (Liu, et al., 2007) proposed two methods: a regression model and a Kalman filter for the ramps with erroneous link-entrance and link-exit detectors. For the Kalman filter measurement, a linear regression model was developed using the occupancy and count measurements of the link-entrance and link-exit detectors as variables. A study by Vigos et al. (Vigos, Papageorgiou, & Wang, 2006) also employs the Kalman filter to improve the accuracy of the flow-in and flow-out approach by continuously adjusting the system state using the time occupancy measurements from an extra loop detector placed in the middle of the link. This method translates time occupancies into space occupancies using the basic relationship between these two measurements in signalised links as demonstrated by Papageorgiou and Vigos (M. Papageorgiou & Vigos, 2008). The time occupancy measurement is an unbiased estimation and does not accumulate over time. Therefore, the mid-link occupancy can be used to generate a correction (or measurement) term in the Kalman filter framework. This method can produce reliable queue estimations while the queue end is fluctuating around the mid-block detector. However, for long queues where the queue end is located upstream of the mid-link detector, the mid-link occupancy will be constantly high. Thus it is no
longer an unbiased estimation and the generated correction term is no longer an accurate correction term.

### 3.3.2 Algorithm Development

The Kalman Filter, a set of mathematical equations, is an efficient tool to estimate the state of a system by minimising the mean of the squared error (Welch & Bishop, 2001). The filter is very powerful in estimating past, present, and future states even when the precise information of the modelled system is unknown. The Kalman Filter theory has two basic sets of equations including a system state equation and a system measurement equation. The system state equation represents the nature of the system states, and is usually written in the following discrete form:

\[
x(t) = A \cdot x(t - 1) + B \cdot u(t - 1) + w(t - 1)
\]  

where \( x \) is system state vector;
\( A \) is input matrix;
\( u \) is control input vector;
\( B \) is control matrix;
\( w \) is process noise; and,
\( t \) represents the time instance.

The system measurement equation describes the relationship between system states and measurements. Acknowledging that measurements inevitably contain noise, the measurement equation is expressed as follows:

\[
z(t) = H \cdot x(t) + v(t)
\]  

where \( z \) is measurement vector;
\( H \) is output matrix; and,
\( v \) is measurement noise.

**System Estimation and Measurement**

A typical metered motorway on-ramp is illustrated in Figure 3-2. The figure also illustrates the detector requirements for the proposed algorithm. The algorithm estimates the ramp queue size; that is, the number of vehicles between the link-exit detectors and the link-entrance detectors. It assumes three detector sets according to
the typical Australian configuration (Burley; & Gaffney, 2013): link exit, mid-link, and link entrance detectors. The link exit and link entrance detectors provide flow measurements. Occupancy measurements are also required from the mid-link and link entrance detectors.

As noted at the beginning of this section, the algorithm is developed based on the Kalman filter framework. Two linear relationships are assumed to formulate the system state equation and the system measurement equation. The system state equation is formulated using the flow-in and flow-out count differences over time based on the flow conservation law, which can be expressed as follows:

\[ NV(t) = NV(t - 1) + [f_{in}(t) - f_{out}(t)] \]  

(3-3)

where \( NV \) is the system state in terms of the number of vehicles on the ramp;

\( f_{in} \) is the traffic flow-in measured by the ramp entrance detectors; and,

\( f_{out} \) is the traffic flow-out measured by the ramp exit detectors.

The system measurement equation is developed based on the linear relationship between the space occupancy and the number of vehicles (\( NV \)), which can be estimated using the following equation:

\[ NV_{mea}(t) = \frac{O_s(t)l_{ramp}N_{lanes}}{L_{vehicle}100} = NV(t) \]  

(3-4)

where \( O_s \) is the space occupancy;

\( L_{vehicle} \) is the average vehicle length;
\( N_{\text{lanes}} \) is the number of lanes in the on-ramp; and,

\( L_{\text{ramp}} \) is the ramp length.

Space occupancy "\( Q_s \)" is an instantaneous (i.e., at a certain time instance) space extended quantity that reflects the portion of link length covered by vehicles. It is impossible to directly measure the space occupancy using loop detector measurements; alternatively, time occupancies can be converted to approximate space occupancies. Time occupancy is a bias-free estimate of space occupancy in a sufficiently small space-time window, assuming the effective vehicle length equals the physical vehicle length (M. Papageorgiou & Vigos, 2008). The Kalman Filter measurement equation proposed by Vigos et al. (Vigos, et al., 2006) uses the time occupancy from the mid-link detector to update the system state. However, this method has a limitation in congestion conditions when the queue size is constantly long and the time occupancy measurements from the mid-link detector do not represent the actual queue size on the entire ramp.

This algorithm also builds on the relationship between time occupancies and space occupancies; however, it includes the time occupancy from the link-entrance detector in the measurement equation in order to overcome this limitation. The time occupancy measurements from the mid-link and the link entrance detectors are processed to approximate the link space occupancy using the following equation:

\[
\hat{\theta}_s(t) = \begin{cases} 
O_{\text{mid}}(t), & O_{\text{mid}}(t) < O_{\text{con}} \\
50 + \frac{O_{\text{up}}(t)}{2}, & \text{otherwise}
\end{cases}
\]  

(3-5)

where \( O_{\text{mid}} \) is the time occupancy measurement from the mid-link detector;

\( O_{\text{up}} \) is the time occupancy measurement from the link entrance detector;

\( O_{\text{con}} \) is the congestion occupancy when queue iv over the detector; and,

\( \hat{\theta}_s \) is the estimated space occupancy.

The above equation implies that the space occupancy is directly approximated using only the mid-link time occupancy for short queues. For long queues, the algorithm assumes a linear increment of the space occupancy with the increment of the link-entrance time occupancy, at the rate of \( 50 + \frac{O_{\text{up}}(t)}{2} \). In this study, not only
the stationary queue but also the slow moving queue are considered. For the slow moving queue, $O_{con}$ is calculated by the following equation:

$$O_{con} = l_{ef} / (l_{ef} + h_s)$$  \hspace{1cm} (3-6)

where $l_{ef}$ is the average effective length of vehicles; and,

$h_s$ is the queuing spacing.

For the Pacific Motorway (the simulation model used in this study), the heavy-duty vehicle (HV) ratio is less than 5%; therefore, the simulation model considers only passenger car. Accordingly, $l_{ef}$ is set as 4.5 meters while $h_s$ is 2 meters. Therefore, the value of $O_{con}$ is around 70% based on Equation 3-6. Note that the threshold will be different for different HV ratios.

**Kalman Filter Estimator**

The process noise, $w$, is sourced from the loop detector counting errors, and is assumed to be an unbiased error with the mean value of zero. The measurement noise, $v$, is caused when converting the time occupancy into the space occupancy. As the two linear relationships are independent of each other, the two noises are assumed to be independent of each other, and normally distributed with constant variances as expressed follows:

$$p(w) \sim N(0, Q) \text{ and } p(v) \sim N(0, R)$$  \hspace{1cm} (3-7)

where $Q$ is the process noise variance; and,

$R$ is the measurement noise variance.

In the standard Kalman Filter process, a prediction process and a correction (estimation) process are employed. The errors involved in each process are referred to as a priori errors and a posterior error, respectively, and their definitions are as follows:

$$e^-(t) \equiv x(t) - \hat{x}^-(t)$$  \hspace{1cm} (3-8)

$$e(t) \equiv x(t) - \hat{x}(t)$$  \hspace{1cm} (3-9)

where $e^-(t)$ is the priori error;

$\hat{x}^-(t)$ is the prediction;

$e(t)$ is the posterior error; and,
\( \hat{x}(t) \) is the estimation (correction).

Accordingly, the two error covariances are given as follows:

\[
P^{-}(t) = E[e^{-}(t) \cdot e^{-T}(t)] \tag{3-10}
\]

\[
P(t) = E[e(t) \cdot e^{T}(t)] \tag{3-11}
\]

The correction equation is given by:

\[
\hat{x}(t) = \hat{x}^{-}(t) + K(t)[z(t) - H\hat{x}^{-}(t)] \tag{3-12}
\]

where \( K \) is the Kalman gain matrix.

The correction process is to use measurement to correct prediction. The Kalman Filter method attempts to select a \( K \) matrix to minimize the estimation error. One widely used solution for determining the \( K \) matrix is given as follows (Welch & Bishop, 2001):

\[
K(t) = P^{-}(t) \cdot H^{T}(HP^{-}(t)H^{T} + R)^{-1} \tag{3-13}
\]

The purpose of the Kalman Filter is to minimise a posterior estimation error covariance by selecting an appropriate Kalman gain matrix. In other words, the filter uses actual measurements to correct the system state prediction and then obtains a better estimation of the system state. There are two steps at each interval. The first step is a time update, the results from which are used as the prediction. For time update, the prediction is calculated by a system equation and the prediction error covariance is updated:

\[
\hat{x}^{-}(t) = A \cdot \hat{x}(t-1) + B \cdot u(t-1) \tag{3-14}
\]

\[
P^{-}(t) = AP(t-1)A^{T} + Q \tag{3-15}
\]

Note that in this study the Kalman Filter is applied to a single dimensional problem where \( A=1, B=1, \) and \( H=1 \). The next step is the measurement update. For this measurement update, the \( K \) matrix is firstly updated and then the estimation is calculated. Finally, the estimation error covariance is updated as follows:

\[
P(t) = [I - K(t)H] \cdot P^{-}(t) \tag{3-16}
\]

When \( A, H, Q \) and \( R \) are time-independent, the filter becomes stationary after some iterations. In this case, we have \( A=1 \) and \( H=1 \); both the process noise and the measurement noise are from error and from the noise of loop detectors; however,
these noises do not change significantly for a particular loop detector, so both Q and R can be assumed to be time-independent. Accordingly, when the filter has reached its asymptote \( P(t) = P(t+1) = \cdots = P \) and same thing for the other variables), the following relations exist:

\[
K = P^{-}(P^{-} + R)^{-1}
\]

\[
P = (1 - K)P^{-}
\]

\[
P^{-} = P + Q
\]

Then, we can deduce an equation, for \( P^{-} \):

\[
(P^{-})^2 - QP^{-} - RQ = 0
\]

\( P^{-} \) when it is stationary is solved by the above equation (the non-negative root):

\[
P^{-} = \frac{Q+\sqrt{Q^2+4RQ}}{2}
\]

Accordingly, \( K \) and \( P \) are given as follows:

\[
K = \frac{Q+\sqrt{Q^2+4RQ}}{Q+\sqrt{Q^2+4RQ+2R}}
\]

\[
P = R \cdot \frac{Q+\sqrt{Q^2+4RQ}}{Q+\sqrt{Q^2+4RQ+2R}}
\]

The noise ratio, denoted as \( \mu = \frac{Q}{R} \), yields Equation 3-22 as follows:

\[
K = \frac{\mu+\sqrt{\mu^2+4\mu}}{\mu+\sqrt{\mu^2+4\mu}+2}
\]

The value of \( K \) depends on the noise ratio, \( \mu \), rather than on the explicit values of \( Q \) and \( R \). If \( \mu \to 0 \) (i.e. small process noise and significant measurement noise), Equation 3-24 yields \( K=0 \), which indicates no need for correction; on the other hand, \( \mu \to \infty \) yields \( K=1 \), which means that only the measurement is reliable.

**Discussion of Kalman Gain \( K \)**

As \( K \) depends on the noise ratio \( \mu \), the key is to analyse the factors affecting the two noises so as to affect \( \mu \) and selection of \( K \). Firstly, the process noise should be smaller than the measurement noise. This is because the estimation method (i.e., flow-in and flow-out difference) is more accurate and reliable than the measurement method (i.e., converting time occupancies into space occupancies). Therefore, a
small Kalman gain is expected. Secondly, the process noise introduced by detector count errors is only related to the particular detectors, while the measurement noise is related not only to the accuracy of space occupancy estimation but also to the length of the ramp based on Equation 3-4. In order to better understand the space occupancy estimation and its impact on the measurement noise with different ramp length, a simulation analysis is conducted.

In the simulation analysis, a single-lane ramp with three detectors (as shown in Figure 3-2) is controlled by ALINEA algorithm (M Papageorgiou, et al., 1997). Three different ramp length are tested: 120 meters, 200 meters and 300 meters. Assuming no significant error of time occupancy measure from the loop detector, the objective of the analysis is to investigate the impact of ramp length on the accuracy of the space occupancy estimation and the measurement method. Root-mean-square error (RMSE) is used to measure the accuracy. Simulation results are summarised in Table 3-1.

\[
\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\text{Observation} - \text{Estimation})^2}
\]  

(3-25)

<table>
<thead>
<tr>
<th>Ramp length</th>
<th>120 meters</th>
<th>200 meters</th>
<th>300 meters</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space occupancy estimation</td>
<td>18.7</td>
<td>18.1</td>
<td>17.9</td>
</tr>
<tr>
<td>Measured queue length</td>
<td>4.5</td>
<td>7.2</td>
<td>10.8</td>
</tr>
</tbody>
</table>

Table 3-1 Simulation results of the space occupancy estimation and the measured queue length

Overall, the results demonstrate that the space occupancy estimation is stable and independent from ramp length. On the contrast, the measured queue length correlates with ramp length: the longer the ramp is, the lower accuracy the measured queue length has (the higher noise). This implies that the longer the ramp is, the smaller \( K \) should be selected, given other conditions are the same. In order to test this, a simulation test for \( K \) value is conducted for different ramp lengths. The results are demonstrated in Figure 3-3.

It can be seen from Figure 3-3 that different ramp lengths result in different best values of \( K \), and the longer ramp results in a smaller \( K \): the best value for 120-meter long ramp is around 0.25; the best one for 200-meter long ramp is about 0.15; and the best value for 300-meter long ramp is less than 0.1. Accordingly, the
recommendation for Kalman gain $K$ is as follows: 1) select a small value (less than 0.3); and 2) the longer the ramp is, the smaller the gain should be.

![Figure 3-3 Estimation RMSE in dependence of the Kalman gain $K$](image)

Finally, the basic formulation of the queue estimation algorithm can be built as follows:

$$NV_{sys}(t) = NV_{est}(t - 1) + [f_{in}(t) - f_{out}(t)]$$  \hspace{1cm} (3-26)

$$NV_{mea}(t) = \frac{\delta_{spc(t)} \cdot l_{ramp} \cdot N_{lanes}}{l_{vehicle} \cdot 100}$$  \hspace{1cm} (3-27)

$$NV_{est}(t) = NV_{sys}(t) + K \cdot [NV_{mea}(t) - NV_{sys}(t)]$$  \hspace{1cm} (3-28)

where $NV_{sys}$ is the $NV$ calculated using the system equation;

$NV_{mea}$ is the $NV$ calculated using the measurement equation;

$K$ is the Kalman gain;

$NV_{est}$ is the $NV$ estimation, the result of the correction equation.

The correction Equation 3-29 can be re-arranged as follows:

$$NV_{est}(t) = (1 - K) \cdot NV_{sys}(t) + K \cdot NV_{mea}(t)$$  \hspace{1cm} (3-29)

In Equation 3-29, the estimate is a smoothed value of $NV_{sys}$ and $NV_{mea}$. Thus, the selection of the Kalman gain must consider the relative size of the system noise.
and the measurement noise. Since the noise of $NV_{sys}$ is much smaller than the noise of $NV_{mea}$, $K$ must be a small number to make $NV_{sys}$ the dominant term in the smoothing.

**Singular Point Correction**

The idea behind single point detection is that an extra detector at the mid-link position can observe when the queue end passes the detector in either a forward or a backward direction. For instance, a significant increase in the observed occupancy may indicate that the queue end has passed the detector location backward. On the contrary, a rapid reduction in the occupancy value implies that a forward moving queue (i.e., dissipating queue) has passed the detector position. Figure 3-4 shows the occupancy measurements from a mid-link detector in comparison to the actual number of vehicles, based on micro-simulation data.

![Figure 3-4 Mid-link time occupancies and number of vehicles](image)

The yellow circles indicate the occurrence of a significant increase or decrease of the mid-block occupancy value. It is clear in the graph that the detector occupancy changes sharply when the number of vehicles increases or decreases over half of the ramp storage (around 15 vehicles). This phenomenon can be used to reset the estimated number of vehicles. The single correction point is defined as the instances when the occupancy measurements from the mid-link detector drop or spike in a short-time period. A single correction point is defined as follows:

$$\Delta O_{mid} = |O_{mid}(t) - O_{mid}(t - 1)| > \gamma$$  \hspace{1cm} (3-30)
where $\Delta O_{mid}$ is the observed occupancy increment;

$O_{mid}(t)$ is the time occupancy measurement from the mid-link detector in the t-th interval;

$\gamma$ is the single point correction threshold.

This study defines the single point correction threshold $\gamma$ at 35% based on preliminary simulation tests. Therefore, an increment or decrement of the time occupancy greater than 35% will activate the single point correction. Once the single point correction is activated, the estimated number of vehicles in this interval, $N_{est}(k)$, is set at half of the maximum queue size, $0.5 \cdot N_{max}$. This process can effectively eliminate the previously accumulated counting errors.

**Algorithm Flowchart**

With the single point correction term, the proposed queue estimation algorithm is displayed in Figure 3-5.

![Figure 3-5 Queue estimation algorithm flow](image)

The single point correction requires occupancy measurements from the mid-link detector. Detection of a single point will yield the estimated queue size at the set point of $0.5 \cdot N_{max}$ without the Kalman filter processing. The queue estimation algorithm takes the input of traffic count measurements from the link exit and link.
entrance detectors for the system state equation. The occupancy from the mid-link and the link entrance detectors are used for the system state estimation.

### 3.3.3 Algorithm Evaluation

#### Simulation Test-bed and Scenario

The estimation accuracy and reliability of the new queue estimation algorithm were evaluated using the AIMSUN micro-simulation model. The evaluation was conducted on three on-ramps on the Pacific Motorway: the Birdwood Road northbound ramp, the Marquis Street southbound ramp, and the Logan Road northbound ramp. The Pacific Motorway, approximately a 30-km section from Logan City to the Brisbane CBD, is the major commuting route for commuters from south suburbs to the Brisbane CBD. Recurring congestion occurs to the northbound in the morning peak-hours and to the southbound in the afternoon peak-hours. The simulation network used in this study was edited by Queensland Department of Transport & Main Roads, and model parameters calibrated by Smart Transport Research Centre (Chung, Rahman, Bevrani, & Jiang, 2011). Figure 3-6 shows the road geometry of these ramps, including the position of the existing ramp signals. The motorway at both upstream and downstream of merges is three lanes and the on-ramps are two lanes, as illustrated.

![Figure 3-6 Road geometry of the selected on-ramps for the algorithm evaluation](image)

The ramp and mainline traffic volumes were drawn from the actual loop detectors data on 25 May 2010. The simulation period is 5 hours starting from 5 to 10 am, for the Birdwood Road ramp and the Logan Road ramp to mimic the traffic conditions in the morning peak hours. The afternoon peak period was modelled for
the Marquis Street ramp from 2 to 7 pm. Figure 3-7 presents the traffic demand profiles of the selected ramps during the simulation period. The models are calibrated (Chung, et al., 2011), and the time-step of the simulation was 0.45 seconds and the reaction time was 0.9 seconds.

The proposed algorithm was implemented using the application programming interface (API) functions provided by AIMSUN. The algorithm receives the time occupancies and traffic count data (detector aggregation interval is 1-minute) from the simulation model to estimate the space occupancy and queue size based on the proposed method. The processed results were compared against the actual queue and space occupancies observed from the simulation model.

To evaluate the algorithm under realistic conditions, the ALINEA is implemented on the simulation models to reproduce the traffic flow characteristics under RM control. Also, the simple queue flush activates when the occupancy from the link-entrance detector is greater than a threshold value of 70%. The queue flush extends the metering rate to the maximum setting at 900 veh/h/lane for a predetermined intervals to clear off the queue.

In the simulation environment, all the detector measurements have no noise and are perfectly accurate. In order to test the proposed algorithm in a realistic environment, a clipping Gaussian noise in the counting error, ranging between -10% and 10%, was artificially added. The reasons are 1) no pre-knowledge of the detector
accuracy from the field is given, and 2) the average count accuracy of the loop
detector is around 90%. This artificial error is applied to each 1 minute aggregated
count measurement.

**Benchmark Algorithm**

A benchmark algorithm was modelled and evaluated for comparison with the
performance of the proposed queue estimation algorithm. This method mimics the
queue estimation method, utilising only the mid-link time occupancy (Vigos, et al.,
2006). The benchmark algorithm does not use the link entrance occupancy data. A
potential drawback is therefore the estimation reliability in long queue conditions.
The benchmark algorithm is referred to as the single occupancy based Kalman filter
(SOKF) method in the rest of this section.

**Performance Indicator**

The primary measure of the algorithm performance is the comparison of the
actual queue lengths with the estimated values. The study uses three measures of
performance for the model calibration and validation: mean absolute error (MAE),
root-mean-square error (RMSE, see Equation 3-25), and mean percentage error
(MPE). MAE is a measure of the estimation accuracy: a small MAE value indicates
more accurate estimation. RMSE is a measure of estimation stability in number of
vehicles: a smaller RMSE value indicates a higher degree of estimation reliability.
MPE indicates the degree estimation errors in a way relative to the actual queue
length.

\[
MAE = \frac{1}{n} \sum_{n} \left| Observation - Estimation \right| \quad (3-31)
\]

\[
MPE = \frac{\frac{MAE}{\sum_{n} Observation} \times 100\%}{(3-32)}
\]

**Simulation Results and Discussion**

The simulation results are presented and discussed in this section. The queue
estimation performance is presented in comparison to the SOKF method, with and
without singular point correction. A total of twenty replications were performed to
collect the results for each test scenario. Table 3-2 provides the summary of the
simulation results. Note that the percentage in the bracket represents the relative
changes from the benchmark algorithm.
Overall, the new algorithm demonstrated reliable queue estimation performances at all the test sites, outperforming the benchmark algorithm. The observed improvements over SOKF are 60.7%, 62.9%, and 61.6% on average in terms of MAE, MPE, and RMSE, respectively, with the singular point correction enabled. Without singular point correction, the improvements are slightly smaller: 56.4%, 61.4%, and 55.8% in terms of MAE, MPE, and RMSE, respectively.

Singular point correction played a supplementary role to further improve the queue estimation, as expected. In high volume ramps, the ramp queue is likely to stay beyond the mid-block detector position for most of the simulation period. As a result, the impact of singular point correction is insignificant; for example, only 7 activations over the 5-hour period on the Birdwood Road on-ramp.

### Table 3-2 Queue Estimation Algorithm Evaluation Results

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Performance measure</th>
<th>Test on-ramps</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Birdwood Rd</td>
<td>Marquis St</td>
</tr>
<tr>
<td>Benchmark algorithm</td>
<td>MAE</td>
<td>18.45</td>
<td>16.95</td>
</tr>
<tr>
<td></td>
<td>MPE</td>
<td>38.79%</td>
<td>30.86%</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>26.90</td>
<td>22.09</td>
</tr>
<tr>
<td>Proposed algorithm without singular point correction</td>
<td>MAE</td>
<td>6.75 (-63.4%)</td>
<td>5.96 (-64.8%)</td>
</tr>
<tr>
<td></td>
<td>MPE</td>
<td>13.83% (-64.3%)</td>
<td>10.63% (-65.6%)</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>9.63 (-64.2%)</td>
<td>7.61 (-65.6%)</td>
</tr>
<tr>
<td>Proposed algorithm with singular point correction</td>
<td>MAE</td>
<td>6.06 (-67.2%)</td>
<td>5.13 (-69.7%)</td>
</tr>
<tr>
<td></td>
<td>MPE</td>
<td>12.81% (-67.0%)</td>
<td>9.32% (-69.8%)</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>8.98 (-66.6%)</td>
<td>6.62 (-70.0%)</td>
</tr>
</tbody>
</table>

**Birdwood Road ramp**

The Birdwood Road ramp has heavy mainline and ramp volumes during the peak-hours. The ramp is heavily congested from 7:30 to 9:30 am and long queues and queue spillover are frequently observed as a result. The proposed algorithm produced the averaged MAE, MPE, and RMSE at 6.06 (veh), 12.81%, and 8.98 (veh), respectively, which are 67.2%, 67.0%, and 66.6% improvements over the benchmark algorithm.
Figure 3-8 displays the estimated queue sizes by the SOKF method and by the proposed algorithm with singular point correction, in comparison to the actual queue size. The graphs were drawn from one simulation replication as an example. It is clear in the graph that the SOKF method continuously overestimates the queue length, especially in the peak-hours from 7:30 to 9:30 am when long queues are present. The proposed algorithm captures the changes in the queue length reasonably well and the actual queue size does not affect the estimation accuracy. Note that similar results are found from other replications.

![Figure 3-8 Queue estimation results example (Birdwood Road ramp)](image)

**Marquis Street Ramp**

The ramp traffic volume on the Marquis Street ramp is one of the heaviest in the Brisbane area. As a result, the afternoon peak congestion begins at around 3:00 pm and lasts until the end of the simulation period at 7:00 pm. The queue size is constantly long on this ramp, except in the first one hour.

In Figure 3-9, both queue estimation algorithms perform reasonably well until 4:00 pm and then produce overestimations for approximately the next 20 minutes until about 4:20 pm. During this period, the ramp queue size fluctuates significantly, in the range of 80 and 100 vehicles, caused by the queue flush operations. After 4:30 pm, the SOKF method constantly overestimates the queue size whereas the estimates by the proposed algorithm follow the actual queue size quite closely. The actual queue size in this time period fluctuates at around the mid-link position, where the singular point correction activates and supplements the Kalman Filter process.
The simulation results of the Logan Road ramp are similar to the other two ramps. However, the performance difference between the SOKF method and the proposed algorithm is relatively insignificant compared with the other two ramps. The mainline traffic volume is much lower at this location. Although the ramp traffic volume is still significantly high, the moderate traffic condition on the mainline allows less restrictive RM and thus prevents less long queues on the ramp.

Relatively moderate traffic conditions and queue sizes impact on the estimation performance, as indicated in Figure 3-10. The estimation accuracy of the SOKF
method is much improved on those of the other two ramps, although overestimations are still observed between 7:10 and 7:40 am and between 8:40 and 9:00 am when long queues are present. However, the proposed algorithm captures the actual queue size reasonably well during the entire simulation period. The benefit of the innovative concepts of the proposed algorithm is well justified by the improved estimation performance in the high ramp traffic demand conditions, where long queues significantly affect the benchmark algorithm’s performance.

3.3.4 Summary

The algorithm introduces two innovative concepts. Firstly, it continuously corrects the system state estimation using the time occupancy measurements from the mid-link and link entrance detectors. Having the additional occupancy term in the state measurement equation overcomes the limitation of the existing method that the space occupancy estimation could be significantly compromised under heavy ramp traffic conditions. Additionally, a novel singular point correction method is proposed to improve the queue estimation reliability. Although, the singular point correction may occur occasionally depending on the ramp volumes, it can potentially eliminate significant counting errors that accumulate over time and thus substantially improve the queue estimation.

In the performance evaluation, the proposed algorithm demonstrated accurate and reliable estimation performances, constantly outperforming the SOKF method. The observed improvements over the SOKF method are 62% and 63% in average in terms of the estimation accuracy (MAE) and reliability (RMSE), respectively. The singular point correction feature played a supplementary role to further improve the queue estimation. The benefit of the innovative concepts of the algorithm is well justified by the improved estimation performance during peak-hours, when long queues are present. The proposed algorithm captured the actual queue size reasonably well in peak-hours when the performance of the benchmark algorithm significantly compromised.
3.4 QUEUE MANAGEMENT SCHEME

In this section, a queue management scheme is proposed based on real-time queue information provided by the above queue estimation algorithm. The queue management scheme includes three major components: the queue control algorithm to calculate metering rate for queue management, using current queue information; the conditions needed for activation / deactivation of queue management; and the mechanism used to determine final metering rate from ALINEA algorithm and queue management.

3.4.1 Queue Control Algorithm

In order to reduce on-ramp queues, queue management tends to increase the basic metering rate from the ALINEA algorithm. This would be calculated based on queue estimation results. The first step is to calculate the change of queue length based on queue estimation; that is, the queue dynamic equation:

$$ NV^-(t + 1) = NV_{est}(t) + T[q_{in}^-(t + 1) - r_{ALINEA}(t + 1)] $$  \hspace{1cm} (3-33)

where NV is the number of vehicles between the ramp entrance detector and the stop-line detector;

subscript “est” means that it is an estimation term;

super minus “^-” means that it is a prediction term;

$NV_{est}(t)$ is the estimated NV in the current time interval by the queue estimation algorithm;

$q_{in}^-(t + 1)$ is the forecasted new vehicle arrivals for the next time interval;

$r_{ALINEA}(t + 1)$ is the basic metering rate calculated by the ALINEA algorithm, based on Equation 2-2.

The queue control algorithm is designed to calculate an increment on top of the basic metering rate. The amount of increment is determined by the following equation.

$$ \Delta r'(t + 1) = \Delta r(t) + K[NV^-(t + 1) - NV_{target}] $$  \hspace{1cm} (3-34)

where $\Delta r'(t + 1)$ is the candidate metering rate increment for the next time


interval \( (t + 1) \);

\( \Delta r(t) \) is the metering rate increment applied in the current interval;

\( NV_{target} \) is the target NV \((0.7NV_{max})\);

\( K \) is a coefficient converting the number of overflow vehicles \([NV^{-}(t + 1) - NV_{target}]\) to a metering rate. The coefficient value is set at 20 in this study.

According to Equation 3-34, the calculated increments may increase exponentially if the queue size \( (NV) \) does not reduce under the target \( NV \). In addition, the increment applies to the basic metering rate, only when the calculated results are positive. In other words, queue management algorithm only increases the basic metering rate. This can be expressed in an equation as follows:

\[
\Delta r(t + 1) = \max[0, \Delta r'(t + 1)]
\]

where \( \Delta r(t + 1) \) is the metering rate increment that applies to the basic metering rate for the next time interval \((t+1)\).

The algorithm can effectively reduce the risk of queue spillover. However, when the onset of queue spillover is detected, the queue management algorithm amends the queue management scheme to queue flush. Therefore, the metering rate for queue management is given as follows:

\[
r_{QM}(t + 1) = \begin{cases} 
  r_{ALINEA}(t + 1) + \Delta r(t + 1), & \text{if } Occ_{up}(t) < 70 \\
  r_{max}, & \text{otherwise queue flush}
\end{cases}
\]

where \( r_{QM}(t + 1) \) is the queue management metering rate for next interval;

\( Occ_{up} \) is the time occupancy measurement from the ramp entrance detector;

\( r_{max} \) is the maximum metering rate.

### 3.4.2 Conditions for Activation of Queue Management

Obviously, queue management should be operated only when on-ramp queue length is becoming critical. Therefore, a simple rule is required for determination of activating the queue management strategy.
The outcomes of Equation 3-33 are the estimated queue size in the next interval. Activation and deactivation of queue management is determined by comparing those computation results, NV^−(t + 1) and NV_{est}(t), with pre-defined thresholds in this research:

- For activation of queue management, NV^−(t + 1) must be greater than 0.6NV_{max};
- For deactivation of queue management, NV_{est}(t) must be less than 0.4NV_{max}, and NV^−(t + 1) must be less than 0.6NV_{max}.

### 3.4.3 Determination of Local RM Rate

When applying RM actions, long on-ramp queues are most likely created by restricted metering rate and relatively high arriving traffic flow. Under such conditions, it is appropriate to assume that both mainline and ramp demands are at high levels. Given dense mainline traffic and high ramp demand occur at the same time, RM actions and queue management strategies are actually against each other. With high mainline volumes, RM algorithms, such as ALINEA, require the limiting of ramp traffic enter so that no breakdown happens at the merge area; with the creation of long on-ramp queues, queue management strategies need an increased metering rate to release long queues. When the on-ramp queue is becoming critical, previous queue management strategies just simply override RM operations. This way of operating queue management may cause immediate flow breakdown and may further restrict the access of on-ramp traffic to the mainline as a result. Therefore, the question is how to balance between RM and queue management.

A concept of mainline speed recovery is introduced with the principle of suspending queue management temporarily when the mainline traffic condition deteriorates. Once evidence of potential breakdown is identified, this new strategy switches the RM control back to the basic metering control without queue management. Doing this is trying to recover the mainline traffic condition quickly, and saving the benefits of RM. At the same time, the on-ramp queue spillover is considered as the recovery cost, and a pre-defined period is set at the highest affordable recovery cost threshold. Once the total spillover time has reached or exceeded the pre-defined threshold, the suspension of queue management is over written for considering surface traffic. This is because when the spillover threshold is
reached, it is of high possibility that mainline free flow traffic cannot be maintained without seriously interfering with surface traffic. Usually, this is the peak traffic condition in which both motorways and arterial roads are experiencing huge traffic volumes, and so congestion is unavoidable.

In this research, the queue management suspension becomes effective when the mainline average vehicle speed is less than 45 km/hour. The suspension time setting is 300 seconds. The process flow is demonstrated in Figure 3-11.

![Figure 3-11 Local RM rate determination process](image)

### 3.4.4 Simulation Evaluation

The simulation environment, test-beds and demand scenario are exactly the same as in Section 3.3.3. The proposed scheme was implemented using the application programming interface (API) functions provided by AIMSUN. The API receives the mainline speed, time occupancies and traffic count data from the simulation model, and then calculates local RM rate that is sent back to the simulation model.

**Test Scenario**

In order to comprehensively understand the interaction of RM and queue management, and to demonstrate the benefits of the proposed queue management scheme, four scenarios are tested, which are listed as follows:

- Base case: without RM and serious mainline congestion expected;
- Basic RM: ALINEA algorithm only without queue management;
- Queue flush (QF): the ALINEA algorithm with a simple queue flush type queue management;
• Queue management strategy (QM): the basic ALINEA module with the proposed queue management scheme.

**Performance Indicator**

The four performance measures used in this study are listed below:

• Mainline speed: measures the mainline traffic condition in order to evaluate the metering benefit; the average speed is calculated for the whole mainline section.

• Mainline traffic travel time: the travel time for mainline traffic volumes in the network, which starts from mainline entrance and ends at the network exit. It is calculated from individual vehicle travel time.

• Ramp traffic travel time: measured from the entrance of the on-ramp to the end of the network for only the on-ramp traffic. It is also calculated from individual vehicle travel time.

• Queue spillover time: the total time when ramp queue spills over to upstream arterials. In this study, the total queue spill-back time is defined as the total time when 1-min time occupancy of the ramp entrance detector is over 70%.

**Simulation Results and Discussions**

The simulation results of the proposed queue management scheme are presented in comparison to other strategies. Note that a total of twenty replications are performed to collect the results for each test scenario. A summary of the simulation results and major findings follows, with discussion.

*Birdwood Road ramp*

The simulation results of the Birdwood Road ramp are illustrated in Table 3-3. The simple ALINEA strategy demonstrated that it can improve the mainline traffic flow but with significant disadvantages to the ramp traffic. Since this strategy adjusts the metering rate with consideration of only the mainline condition, ALINEA essentially increases the ramp traffic travel time. As a result, the mainline traffic speed improves from 86.4 to 93.0 km/h or by 7.6% and the mainline travel time decreases from 202.63 to 155.45 seconds or by 23.3%. The trade-off is, however, significant to the ramp traffic. The ramp traffic travel time increases by 278% from
124.03 to 468.78 seconds per vehicle. The total queue spillover period also increases significantly, from zero to 141.75 minutes.

Table 3-3 Queue management evaluation – Birdwood Road ramp

<table>
<thead>
<tr>
<th>Performance Indicators</th>
<th>RM strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Base case</td>
</tr>
<tr>
<td>Mainline Speed (km/h)</td>
<td>80.05</td>
</tr>
<tr>
<td>Mainline traffic travel time (sec/veh)</td>
<td>202.63</td>
</tr>
<tr>
<td>Ramp traffic travel time (sec/veh)</td>
<td>124.03</td>
</tr>
<tr>
<td>Queue spillover period (min)</td>
<td>0.00</td>
</tr>
</tbody>
</table>

This limitation of ALINEA is mitigated to some extent by employing a simple queue management scheme. With the queue flush strategy, the observed ramp travel time is 400.34 seconds per vehicle and the queue spillover period is 55.3 minutes. Those results are 14.6% and 61.0% reductions in the ramp travel time and the spillover period, respectively, compared with the ALINEA strategy. However, the RM benefit reduces with the queue flush. The mainline traffic speed is observed at 87.53 km/h; that is, a 9.3% improvement over the base case, but a 5.6% reduction from the simple ALINEA strategy.

The proposed queue management scheme demonstrated a balanced performance between keeping the benefit to the mainline traffic and reducing excessive delays to the ramp traffic. Against the base case, the mainline traffic speed improves by 11.6% from 80.05 to 89.35 km/h and the mainline travel time decreases by 19.3% from 202.63 to 163.48 seconds per vehicle. This algorithm also effectively prevents excessive ramp delay and queue spillover, compared with other RM strategies. The observed queue spillover period is only 7.8 minutes, which is a significant reduction from 141.75 minutes and 55.3 minutes with ALINEA and the queue flush strategies, respectively.

*Marquis Street ramp*

The simulation results of the Marquis Street on-ramp is provided in Table 3-4. Overall, the evaluation results are very similar to those of the Birdwood Road ramp, but the mainline speed improvements from the RM are more significant. This is because the mainline volume is higher than at the Birdwood merging area. Therefore, regulating ramp traffic results in more positive impacts on the mainline traffic. The observed mainline speed improvement with simple ALINEA is notable at 56.5%, but
the disadvantages to the ramp traffic are also considerable. The ramp travel time and the queue spillover period increase by 281% and 270%, respectively, compared with the base case.

Table 3-4 Queue management evaluation – Marquis Street ramp

<table>
<thead>
<tr>
<th>Performance Indicators</th>
<th>RM strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Base case</td>
</tr>
<tr>
<td>Mainline Speed (km/h)</td>
<td>56.11</td>
</tr>
<tr>
<td>Mainline traffic travel time (sec/veh)</td>
<td>238.87</td>
</tr>
<tr>
<td>Ramp traffic travel time (sec/veh)</td>
<td>143.06</td>
</tr>
<tr>
<td>Queue spillover period (min)</td>
<td>0.00</td>
</tr>
</tbody>
</table>

The QF strategy alleviates the disadvantage of RM to some degree. In comparison to ALINEA, the ramp travel time reduces by 19.8% from 545.06 to 437.13 seconds. The queue spillover period also significantly reduces by 52.5% from 270.25 to 128.5 minutes.

The QM strategy demonstrates reasonable RM benefits to the mainline traffic while minimising the adverse effects to the on-ramp compared with other strategies. The advanced queue management enables more efficient treatment of long queues over the queue flush. The queue spillover period reduces significantly, compared with the queue flush strategy by 42.8% from 128.5 to 73.55 minutes. Simultaneously, the observed mainline speed improves over the queue flush strategy by 2.6% from 61.78 to 63.38 km/h.

Logan Road ramp

Table 3-5 presents the simulation results of the Logan Road ramp. The mainline and on-ramp volumes of the Logan Road on-ramp are relatively low. However, due to the high occupancy vehicle lane (the T2 lane), only two lanes are usable in the mainline for normal (non-high occupancy) traffic.

As a result of these conditions, the impact of RM is more substantial compared to the other test ramps. The mainline condition notably improves with RM. For instance, the proposed strategy improves the mainline speed by 59.3% from 50.8 to 80.95 km/h and also reduces the mainline travel time by 56.1% from 207.38 to 91.12 seconds per vehicle. It also managed the ramp queue reasonably well, compared with other strategies. The observed queue spillover period is only 19.95 minutes out of a
5-hour simulation period. Although simple ALINEA demonstrated slightly better results in terms of the mainline performance over the other strategies, it produced the longest queue spillover period and the highest ramp travel time, at 56.1 minutes and 294.84 seconds/vehicle, respectively. The QF strategy managed long on-ramp queues better than simple ALINEA managed them; however, it is certainly outperformed by the proposed strategy in terms of the mainline and the ramp traffic performances.

Table 3-5 Queue management evaluation – Logan Road ramp

<table>
<thead>
<tr>
<th>Performance Indicators</th>
<th>RM strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Base case</td>
</tr>
<tr>
<td>Mainline Speed (km/h)</td>
<td>50.80</td>
</tr>
<tr>
<td>Mainline traffic travel time (sec/veh)</td>
<td>207.38</td>
</tr>
<tr>
<td>Ramp traffic travel time (sec/veh)</td>
<td>184.55</td>
</tr>
<tr>
<td>Queue spillover period (min)</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Metering Control Process Comparison**

To provide further insight into the difference between the proposed queue management scheme and the QF strategy, Figure 3-12 compares the metering operation by the two strategies at the Birdwood Road ramp as an example. The figure also illustrates the merging area density and the queue size measurements. The metering rate would be adjusted lower when the merging density increases. The proposed scheme would increase the metering rate gradually when the on-ramp queue size increases. However, the QF strategy simply adjusts the metering to the maximum setting when a long queue is detected. Note that the figure shows the metering operation only after 7:30 am, when traffic queues start to build up in the on-ramp.

In the graphs, the proposed scheme certainly manages the ramp queue more effectively and efficiently than the QF strategy. The ramp queue is consistently managed lower with the QM scheme, which assists the metering to operate without queue flush and which consequently prevents the mainline congestion. In the figure, the queue begins to lengthen after around 7:45 am and the QM scheme reacts to the queue by gradually increasing the metering. This operation maintained relatively high metering rates between 8:00 and 8:15 am, which was effective in preventing abrupt metering increases (i.e., queue flush). As a result, it prevented the congestion between 8:00 and 8:15 am, which had occurred with the QF strategy.
3.4.5 Summary

This section presents a local queue management scheme; a new concept of mainline speed recovery is proposed to balance the conflict between RM and queue management. Microscopic simulation results suggest that the smart queue management scheme can achieve a good balance between mainline and on-ramp performances.
3.5 SUMMARY

Two critical components for a local RM sub-system with advanced queue management have been developed. One is the queue estimation algorithm in real time: micro-simulation results indicate high accuracy of the queue estimation algorithm. The other one is the queue management scheme with smart balance between RM and queue management: micro-simulation results show clearly its well-balanced metering control.

With these two components, the local RM sub-system in this research is established: ALINEA is the basic RM algorithm, and a smart queue management with real-time queue estimation works on top of it.
Chapter 4  Ramp Coordination

4.1 INTRODUCTION

The local RM sub-system was established in Chapter 3, and the next step for establishing the bench mark is to build the coordination between on-ramps. The chapter presents the development of the coordination algorithm between on-ramps for constructing the bench mark for recovery study.

The main reason for requiring ramp coordination is the unbalanced traffic distribution along motorway networks. Specifically, traffic is unevenly distributed along the whole network. Heavy ramp flows and lane reductions are the main cause of recurrent congestion of the motorway network. With only local ramp metering (LRM), it is impossible to handle those on-ramps with heavy traffic flow due to queue constraints. As a result, traffic queues developed in those areas propagate upstream, activating other bottlenecks. In addition, only the on-ramps located close to the bottleneck would take action restricting the mainline access, given that localised RM independently is operated. Meanwhile, upstream ramps are not efficiently utilised because they would not detect congestion from their local information. To sum up, ramp coordination is required for better utilisation of network resources for congestion management.

In Chapter 2, the rule-based heuristic approach has been already identified as the current state of practice. Accordingly, the new coordination algorithm takes the rule-based heuristic approach, with the feedback concept embedded in the control structure. The proposed control approach is simple, transparent and less data-dependent.

Considering the limited roadway infrastructures and the excessive traffic demands, it is infeasible to completely eliminate traffic congestion with only RM. Consequently, the primary objective of the coordinated RM is defined as follows:

*The coordinated RM control aims to delay the onset of congestion and to share the “pain” (on-ramp delay and long queues) among all the metered on-ramps by adopting coordination.*
The next section presents the framework of the ramp coordination. In Section 4.3, the most important component, the slave PID (Proportional–Integral–Derivative) controller, is developed for incorporating the feedback concept. The simulation evaluation results and discussion appeal in Section 4.4. Section 4.5 summaries this chapter with several conclusions. For convenience, the coordinated RM algorithm developed in this study is named the CRM.

4.2 MULTILAYER CONTROL FRAMEWORK

4.2.1 The Framework

To effectively delay the onset of congestion, coordination must activate in advance of flow breakdown. This requires the coordination control to be proactive. A proactive control needs the system information not only for the current but also for the near future. This essentially generates errors and prevents the control output from being the optimal solution. To overcome this limitation, the CRM incorporates the concept of reactive control to adjust new control variables (i.e., the metering rates) based on observed traffic states. A multi-layer control framework is suggested to combine and take advantages of both the proactive and reactive approaches.

Figure 4-1 illustrates the basic framework of the CRM and the main control elements in each layer.

![Figure 4-1 Multilayer control framework](image)

The higher level layer (or coordination control layer) is a centralised, predictive controller that activates the coordination control when the traffic measurements indicate an imminent flow breakdown in the near term future. Coordination is realised by restricting the mainline access at one or more metered ramps in the
upstream of the potential bottleneck area. Another important task undertaken at this layer is to dynamically define the coordination group based on the prevailing traffic condition at active bottlenecks.

The lower level layer (or slave control layer) incorporates reactive controllers that determine the metering rates of those ramps in the coordination group. The slave metering rate is calculated based on both the traffic density (loop detector occupancy) level in the downstream bottleneck area and its own ramp queue size. The control mechanism is a feedback approach that adjusts the slave metering rate continuously to achieve the desired traffic condition in the downstream bottleneck area.

The difference between the two layers is the update interval. In particular, the update interval at the coordination control layer is longer than the counterpart at the slave control layer. This is because the coordination control layer constructs the coordination affecting a large part of motorway network (could be up to about a 10-km section) with predictive information. Therefore, the effects have delays from upstream on-ramps to the downstream bottleneck. At the slave control layer, the slave controller considers both the actual changes at the downstream bottleneck and the slave’s own condition, so a shorter update interval is adopted to enable quick feedback reaction on time.

### 4.2.2 Flowchart

The CRM algorithm consists of five components, as displayed in Figure 4-2; the flowchart of control is also presented. The first step is to identify the ramp(s) in need of coordination (multiple masters allowed at one time). A ramp with the queue size exceeding a certain threshold becomes a “master” ramp that requests coordination. For each master ramp, the slave selection component recruits one or more upstream ramps as “slaves”, switching their metering to the coordinated mode. The coordination group can be resized on a regular basis. A congested slave ramp will be released from the coordination. A new slave ramp could be recruited to replace the released one or to give additional aid to the master ramp. The last component cancels the coordination when the queue size in the master ramp reduces under a pre-specified level or all the available ramps are used up.
The three components in the dashed line area work at the coordination control layer, while the other two are at the slave control layer.

4.2.3 Coordination Activation

This component identifies a master ramp and activates coordination. Three conditions would activate coordination: 1) a mainstream traffic state approaching the merging area capacity; 2) a ramp queue size exceeding a threshold level; and, 3) a ramp queue size projected to spill-over in the near future. The mainstream traffic state is measured and projected using the single exponential smoothing technique. The ramp queue size can be estimated using the on-ramp queue estimation algorithm presented in Section 3.3. For the ramp queue size projection, this study suggests a simple but robust technique to forecast the trend of the queue size change.

Data Smoothing

Data smoothing is used to filter the influence of the randomness in detector measurements and to keep the trend of change. In this study, a single exponential smoothing is chosen for its simplicity, and the process is given as follows:

\[
\begin{align*}
S_0 &= R_0 \\
S_i &= \theta \cdot R_i + (1 - \theta) \cdot S_{i-1}
\end{align*}
\]  

(4-1)

where \(i\) represents the time index;
$S$ is the smoothed data;

$R$ is the raw detector measurement; and,

$\theta$ is the smoothing constant: 0.3 is used in this research as suggested by Queensland Department of Transport and Main Roads (2008).

Figure 4-3 demonstrates the superiority of using smoothed data. In the figure, the x-axis represents time and the y-axis shows the coordination state (i.e., active and non-active). The figure clearly shows that using non-smoothed data causes highly fluctuating coordination, while the operation is more stable using smoothed data.

**Queue Projection**

This method was designed to predict the short-term changes of the ramp queue size. The basic idea is using the recent arrival and departure patterns identified from detector measurements to project them into the future control interval. Figure 4-4 illustrates the projection technique. The difference between two curves is the number of vehicles or the queue size in the ramp. The existing queue size is estimated using the on-ramp queue estimation algorithm presented in Section 3.3. The recent arrival and departure rates can be obtained from the traffic detectors. Projected queue size is estimated by smoothing these recent trends for the next control interval (5 minutes in this research) using the double exponential smoothing technique.

The double exponential smoothing technique is appropriate to capture recent pattern in the measurements and to reflect it in the projection. The smoothing process of the double exponential technique is given below:
\[
\begin{align*}
S_i &= \alpha R_i + (1 - \alpha)(S_{i-1} + G_{i-1}) \\
G_i &= \beta (S_i - S_{i-1}) + (1 - \beta) G_{i-1}
\end{align*}
\] (4-2)

where \( i \) is the time index;

\( S \) is the smoothed data;

\( R \) is the raw detector measurement;

\( G \) is the smoothed trend value;

\( \alpha \) is the smoothing constant; and,

\( \beta \) is the trend-smoothing constant.

The accuracy of double exponential smoothing is tested based on field data. Three on-ramps with ramp signal installed are used for data collection. The three on-ramps are Logan Road ramp, Mains Road ramp and Sports Drive ramp on the Pacific Motorway Northbound, Brisbane, Australia. The raw data is 1-minute vehicle count from the ramp entrance loop detector and the exit loop detector. Root-mean-square error (RMSE), and mean percentage error (MPE) are used to evaluate the accuracy (see in Section 3.3.3). Two different aggregation periods, 3-minute and 5-minute, are tested, and the results are shown in Six combinations of double exponential parameters, \( \alpha \) and \( \beta \), are also tested for calibration. The same performance indicators are used and the results are demonstrated in Table 4-2. From the results, \( \alpha=0.3 \) and \( \beta=0.2 \) are selected.

Table 4-1. Based on the results, 5-minute aggregation is selected because of its lower MPE.
Six combinations of double exponential parameters, $\alpha$ and $\beta$, are also tested for calibration. The same performance indicators are used and the results are demonstrated in Table 4-2. From the results, $\alpha=0.3$ and $\beta=0.2$ are selected.

Table 4-1 Evaluation results of 2 aggregation periods

<table>
<thead>
<tr>
<th>Aggregation interval</th>
<th>3-minute</th>
<th>5-minute</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance indicator</td>
<td>RMSE</td>
<td>MPE</td>
</tr>
<tr>
<td>Logan Road</td>
<td>5.66</td>
<td>14.1%</td>
</tr>
<tr>
<td>Mains Road</td>
<td>8.97</td>
<td>10.9%</td>
</tr>
<tr>
<td>Sports Drive</td>
<td>6.93</td>
<td>12.3%</td>
</tr>
</tbody>
</table>

Table 4-2 Evaluation results for parameter calibration

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0.3</th>
<th>0.2</th>
<th>0.1</th>
<th>0.3</th>
<th>0.3</th>
<th>0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>Logan Road</td>
<td>RMSE</td>
<td>7.48</td>
<td>7.47</td>
<td>7.76</td>
<td>7.36</td>
<td>7.25</td>
</tr>
<tr>
<td>Road</td>
<td>MPE</td>
<td>12.13%</td>
<td>12.10%</td>
<td>12.56%</td>
<td>11.93%</td>
<td>11.69%</td>
</tr>
<tr>
<td>Mains Road</td>
<td>RMSE</td>
<td>11.81</td>
<td>12.77</td>
<td>14.56</td>
<td>11.81</td>
<td>11.81</td>
</tr>
<tr>
<td>Road</td>
<td>MPE</td>
<td>9.54%</td>
<td>10.23%</td>
<td>11.83%</td>
<td>9.46%</td>
<td>9.40%</td>
</tr>
<tr>
<td>Drive</td>
<td>MPE</td>
<td>9.88%</td>
<td>9.65%</td>
<td>11.51%</td>
<td>9.67%</td>
<td>9.52%</td>
</tr>
<tr>
<td>Average</td>
<td>RMSE</td>
<td>9.60</td>
<td>9.91</td>
<td>11.19</td>
<td>9.52</td>
<td>9.46</td>
</tr>
<tr>
<td>MPE</td>
<td>10.52%</td>
<td>10.66%</td>
<td>11.97%</td>
<td>10.35%</td>
<td>10.20%</td>
<td>10.14%</td>
</tr>
</tbody>
</table>

**Conditions to Active Coordination**

The aforementioned three conditions to activate coordination can be formulated as follows. Note that all the three conditions must be satisfied.

\[
\begin{align*}
    &\text{Occ}_{i}^{\text{sm}}(t) > \text{Occ}_{i}^{\text{th}} \\
    &\text{NV}_{i}^{\text{est}}(t) > \text{NV}_{i}^{\text{th1}} \\
    &\text{NV}_{i}^{\text{proj}}(t) > \text{NV}_{i}^{\text{th2}} 
\end{align*}
\]  

(4-3)

where *Occ* is detector occupancy;

*i* is the ramp index starting from the most upstream one to downstream;

“sm” indicates a smoothed value;

“th” indicates a threshold value;
“est” indicates a estimated value;
“prj” indicates a projected value; and,

\( NV \) is the queue length in ramp in terms of the number of vehicles.

**Control Intervals and Minimum Activation Period**

To activate coordination promptly to rapidly changing traffic conditions, the activation component runs in every local control interval, which is one minute in this study. Once a master is identified, the coordination control overrides the normal metering control for the next coordination interval, which is five minutes in this study. The coordination might be cancelled in the next coordination control interval.

**4.2.4 Slave Ramp Selection**

The process of slave selection is to seek sufficient “assistance” from slave ramps for the master ramp. The first step is to estimate the level of assistance required. This “requirement” is for the excessive queue in the master ramp (the X symbol in Figure 4-5) to mitigate through the coordination metering and is defined as the difference between the projected queue size and the maximum acceptable queue size, as displayed in Figure 4-5.

![Figure 4-5 Master requirement](image)

The next step is to calculate the possible “contribution” from upstream ramps. The contribution is calculated for each ramp starting from the immediate upstream ramp of the master ramp until the sum of contributions exceeds the master requirement. In Figure 4-6, the potential contribution is denoted by the symbol, Y,
and is defined as the difference between the maximum acceptable queue size and the projected queue size. Note that the ramp with the projected queue size exceeding the maximum acceptable queue will not be recruited.

![Figure 4-6 Slave contribution](image)

**4.2.5 Coordinated Metering Control**

This module controls the metering rate of all the metered ramps in the coordination group. This is a feedback controller and two strategies are included for master and slave ramps.

In the coordination mode, the master ramp will keep the local metering rate. The slave metering control will be determined by a PID controller.

PID controller is the most widely and successfully used controller type due to its simple and transparent form.

The fundamental concept of the slave metering control is to let slaves react to the master ramp’s traffic condition using the PID controller. The physical meaning of proportional, integral and derivative terms are as follows:

- **P-term**: the character ‘P’ stands for “proportional” and the P-term is designed to react for the instant error between target value and instant measurement. Consequently, it is calculated as the change of the accumulative error at interval t.

- **I-term**: ‘I’ means “integral”, which indicates that the I-term reacts to the accumulative error at current interval.
- D-term: ‘D’ equals “derivative”. Accordingly, the D-term represents the trend of the instant error, and in discrete form it is calculated as the change of the accumulative error change.

A standard discrete PID controller is given as follows:

\[
    r(t) = r(t-1) + K_p \cdot [e(t) - e(t-1)] + K_i \cdot e(t) + K_d \cdot [(e(t) - e(t-1)] - [e(t-1) - e(t-2)]
\]  

where \( r \) represents metering rate;

\( K_p, K_i, K_d \) are the coefficients for P-, I- and D-term; and,

\( e \) represents the error between measurement and desired value, given by:

\[
    e(t) = Occ^* - Occ(t)
\]

where \( Occ^* \) is the pre-defined desired occupancy for the controller; and,

\( Occ(t) \) is the occupancy measurement at interval \( t \).

Noted that the detector occupancy is an aggregated measurement, so the error, \( e(t) \), calculated in the above equation is the accumulative error during interval \( t \).

Finally, the more restrictive metering rate between the local metering (i.e., the local RM sub-system proposed in Chapter 3) and the PID controller is selected for implementation. Accordingly, the slave metering control can be formulated as follows:

\[
    \begin{align*}
    r^c &= PID(Occ^M) \\
    r &= \min(r^c, r^L)
    \end{align*}
\]  

where \( r^c \) is the coordinated metering rate;

\( PID() \) is the PID based slave controller;

\( Occ^M \) represents the measurements from master merge area;

\( r^L \) is the local metering rate; and,

\( r \) is the final metering rate to implement.

**4.2.6 Slave Ramp Release and Additional Recruit**

Although the decision to recruit a slave ramp is made based on the projected queue size, the queue projection is always subject to a forecasting error, so it is
possible to create unacceptably long queues in the slave ramp. The release and recruit operations are only available between two slave ramp selection. Once a slave ramp encounters its own queue problem, it must be released from the coordination and the mode of operation must also switch back to the normal local RM mode. In order to prevent those released ramps from taking benefit from other slaves located upstream, the module sets the maximum metering rate with the arrival flow rate.

When one or more slave ramps are released from coordination, the module will search and recruit additional ramps to replace those released. The recruitment of additional ramps is performed by the ramp recruit module described in Section 4.2.4.

4.2.7 Coordination Cancellation

The coordination control might be cancelled by two conditions. One is that the master ramp is no longer in need of coordination because of enhanced traffic flow conditions. The other condition is that the master merging area falls into congestion so coordination is no longer an effective prevention measure. Either of these two conditions may cancel the coordination control and restore the local RM. These conditions can be formulated as follows:

\[ NV_i^{est}(t) < NV_i^{thd}, Occ_i^{sm}(t) < Occ_i^{thd} \]

or

\[ Occ_i^{sm}(t) > Occ_i^{tha} \]

where \( NV_i^{thd} \) is the deactivation queue length threshold;

\( Occ_i^{thd} \) is the deactivation merge occupancy threshold;

\( Occ_i^{tha} \) is the activation merge occupancy threshold.

4.3 PID CONTROLLER FORMULATION

In a PID controller, each P, I, and D term has a different stress on the control output by the magnitude and consistence of observed errors. The error term is defined as the difference between the observed system condition and the targeted system state. In this section, the effectiveness of the PID controller terms is analysed and their parameters are analysed to define the most appropriate settings of the PID controller. Once the formulation of the slave controller is established, the structure of the slave controller is decided, and the next task is to test for the best parameter group based on simulation.
4.3.1 Analysis of Controller Terms

To better understand the operation of the PID controller, the following experiment was conducted to demonstrate the impact of each term on the control output. Table 4-3 shows the values of P, I and D terms corresponding to gradually increasing occupancies. In Table 4-3, the time interval is 1-minute (63 seconds in the simulation due to the calibrated reaction time). The test mimics a situation where traffic congestion is building up. The calculation results indicate how each term changes the metering rate. For example, a negative value will decrease the metering rate, while a positive value will increase it. A greater term value will make a more significant change to the existing metering rate. Note that the target occupancy is 26% and the term values for the first two intervals are missing because the PID controller requires the most recent two measurements according to Equation 4-4.

Table 4-3 PID terms in RM

<table>
<thead>
<tr>
<th>Time Interval</th>
<th>Merging area occupancy (%)</th>
<th>Error signal</th>
<th>P-term</th>
<th>I-term</th>
<th>D-term</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>10</td>
<td>-2</td>
<td>10</td>
<td>-1</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>7</td>
<td>-3</td>
<td>7</td>
<td>-1</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>3</td>
<td>-4</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>2</td>
<td>-1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>25</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>27</td>
<td>-1</td>
<td>-2</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>9</td>
<td>26</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>29</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
<td>-4</td>
</tr>
<tr>
<td>11</td>
<td>27</td>
<td>-1</td>
<td>2</td>
<td>-1</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>26</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-1</td>
</tr>
</tbody>
</table>

According to the test results, the characteristics of the PID controller can be summarised as follows:

- P-term changes the metering rate directly proportional to the occupancy change. This feature is useful to adjust the slave metering rate when the merging occupancy is gradually increasing but is yet under the target occupancy.
• I-term begins to reduce the metering rate only after the merging area occupancy rises over the target occupancy. This reaction of slave ramps is obviously too late, considering the time lag caused by the travel time between the master ramp and slave ramps. Therefore, the I-term is inappropriate for the slave metering control.

• When the master occupancy consistently changes in either positively or negatively, the D-term accelerates the reaction (i.e., increasing or decreasing the metering rate) of slave ramps. Consequently, the D-term can be used to supplement the P-term to enable a quicker response of slave ramps when the master occupancy is rising quickly.

The PID controller terms were further analysed to compare their relative effectiveness and suitability for the slave metering control. Each controller term was individually tested and the coefficient was set at 70 for a fair comparison. The test was performed on the Pacific Motorway test-bed using the northbound traffic in the morning peak hour scenario (Chung, et al., 2011) (the details of the test-bed are presented in Section 4.4.1). Traffic congestion is heavy along the motorway and merging causes a major bottleneck between the Loganlea Road and Service Road on-ramps. This section was chosen for the test and the metering rates of the master and slave ramps were recorded when congestion is building up. Table 4-4 summarises the simulation results of 10 replications in selected performance measures.

Table 4-4 Analysis of the controller term effectiveness

<table>
<thead>
<tr>
<th></th>
<th>P-term</th>
<th>I-term</th>
<th>D-term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average master metering</td>
<td>805</td>
<td>778</td>
<td>770</td>
</tr>
<tr>
<td>rate (veh/h/lane)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average slave metering</td>
<td>728</td>
<td>685</td>
<td>815</td>
</tr>
<tr>
<td>rate (veh/h/lane)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Next time when calling</td>
<td>75</td>
<td>67</td>
<td>68</td>
</tr>
<tr>
<td>for coordination (interval)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As presented in the table, P-term results are in the highest level of the master metering rate which indicates the best mainstream traffic condition. In terms of the average slave metering rate, the P-term also performs better than the I-term, but the D-term produces the best result. Although the D-term allows more ramp traffic to access the mainline in slave ramps, their contribution to the master ramps is relatively insignificant compared with the P-term. The I-term is relatively ineffective
among other terms. In summary, the simulation results confirm the characteristics of
the PID controller. Firstly, the P-term is the most appropriate measure for the slave
metering control under the traffic condition where a breakdown is anticipated sooner
or later. Secondly, the D-term is also an effective control measure to assist the P-term
time control and to accelerate the response of slave ramps. Thirdly, the I-term is
inappropriate for the slave metering control so it is excluded from the formulation.
Note that the I-term is mandatory for eliminating stationary error of the PID
controller, but it is not mandatory for the slave controller because the main
controller itself is an I-controller (ALINEA).

4.3.2 Analysis of Controller Formulations and Coefficients

This section analyses the PID controller coefficients. A two-step calibration is
performed to find the best coefficient settings for slave ramp groups classified by
their travel time (the distance) to the master ramp. The slave RM control comes into
effect when the ramp traffic from those slave ramps arrives at the master merging
area. This implies that the response speed of slave ramps should differ from their
distance (or travel time) to the master ramp. Therefore, the CRM classifies the
metered ramps located upstream of the master ramp in a few categories by their
travel time to the master ramp, applying different PID controller structures. Four
groups are defined as shown in Table 4-5. The grouping setting presented in Table
4-5 is adjustable.

<table>
<thead>
<tr>
<th>Group</th>
<th>Travel time to the master (assuming 80 km/h as the speed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>1.5 minutes</td>
</tr>
<tr>
<td>Group 2</td>
<td>From 1.5 minutes to 3.5 minutes</td>
</tr>
<tr>
<td>Group 3</td>
<td>From 3.5 minutes to 6 minutes</td>
</tr>
<tr>
<td>Group 4</td>
<td>Over 6 minutes</td>
</tr>
</tbody>
</table>

According to the analysis results of the PID controller, the P-term must be the
main control term, while the D-term is supplementary to accelerate the slave
metering control. Therefore, to find the best structure of PID controller for those four
groups defined in Table 4-5, the following structure forms are tested, as shown in
Table 4-6.
As the last step, each formulation is tested with a range of different coefficient values for P-term and D-term. The tested coefficient range for P-term is between 40 and 140 with an increment of 10. For D-term, the test value ranges between 40 and 120 with an increment of 10. The following performance measures were used to quantify the relative effectiveness of tested controller structures and coefficients.

Table 4-6 Candidate of controller structure

<table>
<thead>
<tr>
<th>Structure</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure 1</td>
<td>P-term</td>
<td>P-term</td>
<td>P-term</td>
<td>P-term</td>
</tr>
<tr>
<td>Structure 2</td>
<td>P-term</td>
<td>P-term</td>
<td>P-term</td>
<td>P- &amp; D-term</td>
</tr>
<tr>
<td>Structure 3</td>
<td>P-term</td>
<td>P-term</td>
<td>P- &amp; D-term</td>
<td>P- &amp; D-term</td>
</tr>
<tr>
<td>Structure 4</td>
<td>P-term</td>
<td>P- &amp; D-term</td>
<td>P- &amp; D-term</td>
<td>P- &amp; D-term</td>
</tr>
<tr>
<td>Structure 5</td>
<td>P- &amp; D-term</td>
<td>P- &amp; D-term</td>
<td>P- &amp; D-term</td>
<td>P- &amp; D-term</td>
</tr>
</tbody>
</table>

- Total Travel Time (TTT): the most widely used efficiency indicator at a system level for RM. It is calculated by summing up all the individual vehicle travel times in the network. The unit of TTT is veh·h.
- Total queue spillover time (TQST): the sum of the total time for each on-ramp when ramp queue spills over to upstream arterials. In this study, the queue spillover is defined as 1-min time occupancy of the ramp entrance detector is over 70% (refer to Equation 3-6).

Each controller-coefficient combination was tested in 10 simulation replications. The aggregated simulation results suggest that the formulation given in Table 4-7 produces the best control output.

Table 4-7 PID controller formulation

<table>
<thead>
<tr>
<th>Structure</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>P=40</td>
<td>P=60</td>
<td>P=80; D=40</td>
<td>P=100; D=70</td>
</tr>
</tbody>
</table>

4.4 ALGORITHM EVALUATION

4.4.1 Simulation Test-bed and Test Scenarios

The modelling platform used in the investigation is AIMSUN 6.1. The Pacific Motorway test-bed model was used for this research. The test-bed network is...
approximately a 30-km section of the northbound (inbound) Pacific Motorway (M3) from Logan City to the Brisbane CBD, as displayed in Figure 4-7. This motorway section serves a large volume of commuter traffic in the morning peak-hours, leading to heavy recurrent congestion.

The M3 has five mainline lanes in the sections where the test network starts and ends (i.e., from the Logan Interchange to the Brisbane CBD), and mostly three mainline lanes overall. There are 16 on-ramps and 17 off-ramps along the network. The traffic volume is approximately 130,000 vehicles per day. This motorway section serves a large volume of commuter traffic in the morning peak hours, leading to heavy recurrent congestion. For these reasons, local authorities consider the M3 to be an ideal motorway to deploy RM to improve efficiency.

The simulation network used in this study was edited by Queensland Dept. of Transport & Main Roads, and model parameters calibrated by Smart Transport Research Centre (Chung, et al., 2011). A complete scenario to depict the real traffic demand on the network was developed in terms of traffic state according to PTDS (Public Transport Data Source) database. The selected case day, 15 March 2010, was a regular business day (Monday) with major educational institutions running, with good weather (no rain) and with no incidents reported. The complete scenario was conducted for a period of 17-hour with time interval of 15 minutes. According to the whole day volume contour, the morning peak period was determined as a 5-hour period from 5am to 10am, when the northbound (inbound) motorway witnessed high levels of recurrent congestion.
A total of three test scenarios were modelled in the AIMSUN simulation network for evaluation of the CRM algorithm.

- Base case scenario assumes no RM control;
- Local RM (LRM) scenario operates a local RM control for all 16 on-ramps along the northbound Pacific Motorway;
- CRM scenario operates the coordinated RM control upon an activation of coordination. Otherwise, ramps will operate local RM. The PID controller employs the structure and coefficients presented in Table 4-7.

4.4.2 Performance Indicators

The RM performance is measured using four indicators to demonstrate the benefits and costs of the CRM, compared with the LRM:

- Total travel time (TTT): the same as presented in Section 4.3.2.
- Average mainline traffic delay (MTD): this indicator gives a sense of the coordination benefit. The northbound Pacific Motorway is divided into 31 sections based on the location of metered ramps. For each section, individual vehicle travel time within the section is collected and aggregated into the average section travel time. The sum of average section travel times is the entire motorway travel time. The free flow travel time for the entire motorway is also calculated assuming 80 km/h as the free flow speed. Finally, MTD is defined as the difference between the actual mainline traffic travel time and the free flow travel time. The unit of this indicator is sec/trip.
- Total queue spillover time (QST): the same as presented in Section 4.3.2.
- Average ramp traffic delay (RTD): the way to calculate RTD is slightly different from MTD. Firstly, the aggregated travel time for each ramp is calculated by collecting individual vehicle travel times in ramp. The ramp travel time is collected from the ramp entrance to the downstream merge area. The free flow speed for this section is assumed at 70 km/h. The delay for each ramp is defined as the difference between the actual ramp travel time and the free flow travel time. To consider that the ramp traffic volume
varies by each location, the average RTD is calculated using the following equation. The unit of this indicator is sec/veh.

\[
\text{RTD} = \frac{\sum_i \text{RTD}_i \cdot Q_i}{\sum_i Q_i}
\]

(4-8)

where \( \text{RTD}_i \) is the ramp traffic delay for ramp \( i \); and,

\( Q_i \) is the total volume of ramp \( i \).

4.4.3 Simulation Results

In Table 4-8, the simulation results are summarised in terms of those four performance indicators.

The simulation results indicate that the base case scenario creates the worst overall traffic condition. The observed total travel time is 14221.8 veh∙h; the mainline traffic delay is also substantially higher than other scenarios, at 911.6 sec/trip. Since no RM is implemented in this scenario, the ramp traffic delay time and the total queue spillover time are the least among the scenarios at 59.6 sec/veh and 220.4 minutes, respectively. The queue spillover in the Base case is because mainline congestion blocks the ramp, causing ramp queuing and spillover.

Installation of the local RM system made significant improvements to the overall traffic performance and the mainstream traffic. The total travel time and the average ramp delay time decreased, with the LRM, by 19.9% (from 14221.8 to 11396.6 veh∙h) and 53.1% (from 911.6 to 427.7 sec/trip). The local metering control, however, negatively affects the ramp travel delay and the total queue spillover time.

Table 4-8 Coordinated ramp metering evaluation results summary

<table>
<thead>
<tr>
<th>Scenario</th>
<th>TTT (veh∙h)</th>
<th>MTD (sec/trip)</th>
<th>RTD (sec/veh)</th>
<th>TQST (minute)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base case</td>
<td>14221.8</td>
<td>911.6</td>
<td>59.6</td>
<td>220.4</td>
</tr>
<tr>
<td>LRM</td>
<td>11396.6</td>
<td>427.7</td>
<td>133.8</td>
<td>620.1</td>
</tr>
<tr>
<td>CRM</td>
<td>11285.7</td>
<td>408.1</td>
<td>153</td>
<td>624.9</td>
</tr>
</tbody>
</table>

Comparison of the LRM and the CRM clearly shows that the coordination control makes the mainstream flow much more quickly. The mainline travel delay decreases with the CRM by 4.6% over the LRM. The coordinated control also improved the overall traffic condition. The observed total travel time is 11285.7
veh·h, which is a 1% reduction over the LRM and 20.7% reduction over the base case scenario. Besides, the ramp traffic delay increases by 14.3% with the coordinated control. Restricting the mainline access at additional metered ramps is a trade-off. It is noteworthy that the total queue spillover time increases only marginally, by 0.8%. This implies that the coordination algorithm can efficiently utilise the queue storage of the slave ramps without causing excessive delays or long queues in those ramps.

Table 4-9 compares the ramp traffic delay time for individual ramps. In general, the ramp delay time increases with the LRM, and it further increases with the CRM due to the nature of local and coordinated RM that restricting the mainline access of ramp traffic. Some exceptional figures are observed in the results of Service Road and Stanley Street. These two ramps are recurrent bottlenecks where a heavy merging traffic causes congestion in the morning peak-hours. A delay reduction in these ramps implies that the coordination algorithm effectively improved the mainstream traffic condition and that more ramp traffic could access to the mainstream as a result.

Some exceptional figures are also observed in the results of the Beenleigh Road, the Grandis Street, and the Murrays Road ramp: where the ramp delay time decreases with the LRM, but it increases again with the CRM. The speed contours presented in Figure 4-8 explain this result. In the figure, those three ramps are severely affected by heavy congestion and long queues generated from the Gateway Motorway interchange in the base case scenario. The LRM effectively reduces the traffic congestion in this area, as shown in the speed contour. As a result of the improved mainstream traffic flow, the delay of ramp traffic also decreases. The ramp delay time increases with the CRM, as the mainline access is more strictly restricted at additional metered ramps.

In Figure 4-8, the size of the traffic queue in the City area clearly decreased with the LRM. As a trade-off, the ramp traffic delays substantially increased in those metered ramps in the City area, including Alice Street and Ann Street ramp. Since the mainline congestion has been reduced significantly, the network also recovers more quickly compared with the base case (the areas in the blue ellipses).
Table 4-9 Comparison of average ramp traffic delay by individual ramps

<table>
<thead>
<tr>
<th>Ramps</th>
<th>Average ramp traffic delay (sec/vehicle)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Base case</td>
</tr>
<tr>
<td>Beenleigh Road</td>
<td>6.1</td>
</tr>
<tr>
<td>Grandis Street</td>
<td>8.8</td>
</tr>
<tr>
<td>Murrays Road</td>
<td>30.8</td>
</tr>
<tr>
<td>Centenary Road</td>
<td>25</td>
</tr>
<tr>
<td>Loganlea Road</td>
<td>15.8</td>
</tr>
<tr>
<td>Service Road</td>
<td>110.6</td>
</tr>
<tr>
<td>Fitzgerald Avenue</td>
<td>47.2</td>
</tr>
<tr>
<td>Sports Drive</td>
<td>429.5</td>
</tr>
<tr>
<td>Logan Road</td>
<td>11.1</td>
</tr>
<tr>
<td>Kessels Road</td>
<td>8.6</td>
</tr>
<tr>
<td>Mains Road</td>
<td>20.3</td>
</tr>
<tr>
<td>Birdwood Road</td>
<td>24.4</td>
</tr>
<tr>
<td>Duke Street</td>
<td>7.2</td>
</tr>
<tr>
<td>Stanley Street</td>
<td>46.6</td>
</tr>
<tr>
<td>Alice Street</td>
<td>19.3</td>
</tr>
<tr>
<td>Ann Street</td>
<td>7.2</td>
</tr>
</tbody>
</table>

Figure 4-8 Motorway mainline speed contour (base case vs. LRM)

Figure 4-9 compares the speed contours of the LRM and the CRM. In general, the graphs show that the coordinated control effectively reduces the congestion at the Birdwood Road ramp and in the Stanley Street ramp area. When comparing the
traffic conditions at the recovery period for the major bottleneck (the areas in the blue ellipse), the CRM does not have any special improvement for the recovery. Note that the locations and sizes of traffic queues vary by simulation replications. The speed contours in Figure 4-8 and Figure 4-9 are samples but all the replications produced similar speed contour patterns.

![Figure 4-9 Motorway mainline speed contour (LRM vs. CRM)](image)

Noted from the results, the Gateway Motorway interchange bottleneck is the most serious bottleneck during the morning peak, as can be seen from all the above speed contours. For this particular bottleneck, high traffic flows take the off-ramp to the Gateway Motorway, thereby causing a large amount of lane-changing in the section just upstream from the off-ramp. This is a weaving bottleneck rather than a merging bottleneck. For the weaving bottleneck, RM is not as effective as for the merging bottleneck. Based on our simulation tests, the congestion cannot be well solved without closure of a high demand on-ramp, such as the Sports Drive on-ramp.
4.5 SUMMARY

This chapter presents a coordinated RM algorithm (the CRM). A special emphasis was placed on the practicality of the algorithm to develop a field implementable strategy. Complex mathematical models and optimisation approaches have been excluded because they require comprehensive and highly reliable traffic detector data, which is often implausible in the real traffic condition. The CRM algorithm takes the rule-based heuristic approach, with the feedback concept embedded in the control structure. The algorithm is simple, transparent, and less data-dependent.

The new algorithm consists of five components that perform unique roles: activation of coordination, recruitment and release of slave ramps, slave metering control, and cancelation of coordination. Those logical components are defined in two control layers – a coordination control layer and a slave control layer. The performance of the CRM was evaluated in simulation against the base case assuming no RM and the LRM scenario employing the local RM system presented in Chapter 3. The simulation results revealed the followings:

- The mainstream traffic flow significantly improved with the CRM. The mainline traffic delay reduced by almost 55% over the base case scenario.
- The CRM was more effective in improving the mainline traffic flow. The mainline vehicle delay time decreased by 4.6% with the CRM over the LRM.
- The improved mainstream traffic flow was achieved by more balanced utilisation of ramp spaces to store traffic queues. Although the ramp delay time increased as a result of the coordination control, the total queue spillover time increased only slightly, 0.8% as compared with the LRM scenario. This indicates the coordination is a good “deal”.

To this point, the bench mark for recovery strategy research, the CRM, has been built and is ready for use.
Chapter 5  Recovery Concept

5.1 INTRODUCTION

Through Chapter 3 and Chapter 4, a coordinated RM system has been established as the benchmark for rapid congestion recovery (RCR) study. The purpose of this chapter is to introduce the recovery concept by answering two questions: 1) what the system benefit of doing RCR is, and 2) whether RM as the solo motorway management tool can accelerate mainline congestion.

In order to answer these questions, this chapter firstly analyses the concept of the recovery phase and gives a detailed definition in Section 5.2. In Section 5.3, the traffic flow dynamics at the recovery phase are then investigated to discover the system benefit of RCR; the basic RM operation for RCR - restrictive metering control (RMC) - is proposed accordingly. Section 5.4 studies the impact of RMC on merging bottleneck throughput at the recovery phase. A simulation investigation on RMC and merging bottleneck throughput is conducted in Section 5.5. Section 5.6 summarises the chapter.

5.2 RECOVERY PHASE

5.2.1 The Definition

The general definition of recovery phase given in Section 1.2.2 is not detailed enough for developing RCR strategy. Moreover, the definition of the recovery phase is critical for RCR strategy because it indicates the timing for activating the strategy.

The general definition given in Section 1.2.2 states the most important characteristic: the time sequence – after mainline congestion. There is no recovery if there is no mainline congestion. Based on this understanding, a more detailed definition can be given from two different aspects: the reason for the mainline congestion and the change of traffic flow properties.

Generally speaking, motorway congestion can be categorised as recurrent and non-recurrent congestion. The first category is usually caused by high demand that exceeds network capacity. A typical instance would be the high frequency of congestion during peak hours, the major focus of this research project. Non-recurrent
congestion is usually caused by incidents, such as road accidents and special event. In this type of congestion, the motorway network capacity is profoundly affected.

The cause for the congestion indicates the time when it will recover. Accordingly, the change of traffic flow properties in this study focuses on the change of traffic demand and capacity. For RM, traffic flows can be divided into mainline traffic flow and ramp traffic flow.

The detailed definition of the recovery phase is as follows:

The recovery phase is the phase when total traffic demands start to reduce (for recurrent congestion, it means that the mainline traffic demand has to reduce and ramp demands do not increase) or when the capacity begins to increase (for example, the capacity will increase once an incident is cleared).

5.2.2 Recovery Scenarios

Two recovery scenarios accord with the above definition:

- Recovery from peak hour congestion: this scenario is mainly related to the recurrent congestion caused by high traffic demand. In the scenario, congestion usually happens at fixed bottleneck locations, and historical information can assist the determination of the recovery phase, because the traffic pattern tends to repeat itself.

- Recovery from incident: this is attached with the congestion caused by incident. Since incidents usually happen randomly, this scenario refers to random bottleneck problems.

5.3 **SYSTEM BENEFIT AT RECOVERY PHASE**

As discussed in the chapter introduction, the way to address the first question is to discover the system benefit of RCR. The best way to investigate the possible benefit is to analyse the differences of the traffic flow dynamics at the recovery phase from the peak hours. As the peak hour congestion scenario is the major focus of this research project, the analysis also focuses on this scenario.
Usually, urban motorways would experience demand reduction in both mainline entrances and on-ramps at the end of peak hours. The reduction leads to two differences in traffic flow dynamics of the recovery phase compared with during peak hours. Firstly, decreased demands indicate no more mainline queuing accumulations. Under this assumption, the earlier the mainline queue can be cleared, the more travel time saving can be achieved. This is because once the mainline queue is cleared the traffic condition can easily maintain a free flow condition at mainline, and maintaining a free flow condition can achieve the highest system efficiency, as well as the most travel time saving. Consequently, mainline traffic would enjoy less delay after the quicker recovery. For ramp traffic, the mainline free flow condition gives more opportunities for them to enter the motorway without causing congestion, thereby reducing delays at ramps. This results a high system travel time saving, as proved theoretically by Papageorgiou and Kotsialos (2002). More importantly, not only mainstream traffic but also ramp traffic is able to benefit.

The second difference is that reduced ramp traffic makes it possible to control the costs of RM, including queue spillover and ramp delays. Specifically, on-ramp queues accumulate slower when there is less incoming ramp traffic, which gives a natural reduction of the costs. In addition, quick ramp traffic discharge can be achieved as the mainline is recovered.

In summary, a proper RCR strategy can satisfy two aspects. One is to achieve system benefits for both mainline traffic and ramp traffic might, and the other one is to manage the RM costs.

In this research project, RM is considered as the motorway management tool to conduct RCR. Consequently, the next question is whether RM can help mainline congestion recovery. At the recovery phase, recovering mainline congestion is actually to clear mainline queues as soon as possible. With RM, the choices are to restrict ramp traffic to a low metering rate or to increase the metering rate for discharging ramp traffic. Operating a restrictive metering rate at the beginning of the recovery phase contributes to mainline queue discharge. According to the literature (Cassidy & Rudjanakanoknad, 2005), field data analysis revealed that restrictively metering an on-ramp can recover the higher discharge flow at a merge and thereby increase the merge capacity. Therefore, the basic RM operation for RCR is to run the most restrictive metering rate, the restrictive metering control (RMC).
5.4 RESTRICTIVE METERING CONTROL AND MERGING BOTTLENECK THROUGHPUT

This section studies the basic RM operation for RCR: that is, the relationship between RMC and merging bottleneck throughput at the recovery phase (i.e., the merging capacity).

One important objective of introducing the RM operation is to break large platoons for better and smoother merging. This is because a large platoon from the ramp would affect mainstream traffic significantly. By breaking a platoon into individuals, the effect from merging vehicles can be reduced. In other words, this implies that reducing the merging vehicle effect on mainline traffic would have positive impacts on merging bottleneck throughput – the merging capacity.

The improvement of the merging capacity by RM is easy to understand. At the merging area, mainline traffic, especially in the leftmost lane, would be disturbed by merging traffic. As a result, some drivers would apply brakes to allow ramp vehicles entering, which might slow the traffic flow in the leftmost lane. As the speed is slow, the possibility of vehicle lane changing increases and the other lanes would be affected. In dense traffic flow, all these brakes and lane-changes at the merging area are the disturbances which originally cause congestion. With RM operation, the effect from merging vehicles can be reduced so that the disturbances can be reduced. Consequently, the merging capacity can be improved.

Field data analyses reported in the literature (Cassidy & Rudjanakanoknad, 2005; Geroliminis, et al., 2011) also supports this conclusion. Zhang and Levinson (2010) analysed the data from the well-known Twin Cities RM experiment in Minnesota, USA, and drew the conclusion that RM actually increases active merging bottleneck capacity (averagely 2% during the pre-queue transition period and 3% of queue discharge flow rates after breakdown). Nikolas Geroliminis et al. (2011) also reported that the total capacity of an active bottleneck (mainline and on-ramp) depends on the ratio of the two flows, and that the capacity is smaller when ramp flows are higher. Cassidy and Rudjanakanoknad (2005) stated that “By means of observation and experiment, we show here that metering an on-ramp can recover the higher discharge flow at a merge and thereby increase the merge capacity.”

Furthermore, RMC allows only a minimum amount of ramp traffic entering. In other words, RMC intentionally minimises the effect of merging traffic on mainline
traffic. The less ramp traffic flow there is, the fewer interferences there are for the mainline traffic. Consequently, mainline traffic is more likely to quickly pass the merging area at a high speed. Given that the recovery phase is the time to apply RMC intentionally, the mainline traffic has already queued at the merging bottleneck. All these queuing vehicles provide enough demand waiting to be flushed. As the interferences are minimised by RMC, the throughput of the merging bottleneck at the recovery phase is maximised. This is a “bonus capacity”, to flush mainline queues at the recovery phase. (Note that even assuming fixed merging throughput, RMC would accelerate the mainline queue discharge because the ramp traffic is reduced.) This analysis indicates that RMC at the recovery phase is able to maximise mainline queue discharge and to recover mainline congestion rapidly.

5.5 SIMULATION INVESTIGATION

As analysed above, RMC increases the merging capacity at the recovery phase and this is a qualitative conclusion. In this section, the merging bottleneck throughput is investigated by micro-simulations with or without RMC.

The objective of the simulation investigation is to quantify the difference in merging bottleneck throughput with different RM operations at the recovery phase.

5.5.1 Simulation Settings

The micro-simulation platform is AIMSUN 6.1, described in Chapter 3 and Chapter 4. In the current simulation model (the Northbound Pacific Motorway), the most serious mainline congestion during the morning peak hours is caused by the Gateway Motorway interchange off-ramp, a weaving bottleneck rather than a merging bottleneck. As mentioned before, RM becomes the most effective for motorway congestion when the major bottleneck is a merging bottleneck. Moreover, the objective of this simulation study is for merging bottleneck throughput at recovery phase. Therefore, two modifications have been made to eliminate the weaving bottleneck:

- Modification 1: the diverge flow at the Gateway Motorway off-ramp is reduced. As a result, the number of lane-changes at the upstream section has been reduced and it is no longer a weaving bottleneck. Another effect is an increase of traffic flow into the downstream, which would make the
start point of the T2-lane (the high occupancy lane) an active bottleneck (lane drop bottleneck). Consequently, another modification is then made.

- Modification 2: all the T2-lanes are set back to normal lanes. This eliminates the aforementioned lane drop bottleneck.

With the two modifications, all the major bottlenecks in the new simulation model are merging bottlenecks, named Test-bed L, an ideal test-bed for RM tests. In the new simulation model, the merging area of the Birdwood Road ramp becomes the most serious bottleneck, causing a long mainline queue during the morning peak hours. Figure 5-1 shows the speed contour of the new simulation model. As can be seen clearly, the Birdwood Road ramp merging bottleneck causes huge mainline queuing, and the Stanley Street ramp is another major bottleneck.

![Figure 5-1 Speed contour of the new simulation model - Modified northbound of the Pacific Motorway](image)

The complexity of traffic flow dynamics in this large network is high. Therefore, another simple test-bed, the Birdwood Road ramp network introduced in Section 3.3.3, is used as Test-bed S. Test-bed L provides a complicated circumstance
that is closer to the real situation, while Test-bed S provides a more controlled environment for investigating RMC’s impact.

This is a before-and-after comparison. In the base case, the ramp is metered by the CRM proposed in Chapter 4 for Test-bed L or the Local RM system presented in Chapter 3 for Test-bed S. In the RMC case, the ramp is running a 20-minute RMC at the end of peak hours. A special group of detectors is placed right at the downstream of the merging area to record 1-minute aggregated vehicle count and speed. The layout is demonstrated in Figure 5-2.

![Detector settings](image)

Figure 5-2 Detector settings

The performance indicators are the throughput and the average speed of each lane during the 20 minutes. In order to reduce the impacts of simulation random seeds, 10 replications are simulated.
5.5.2 Results and Discussion

Table 5-1 shows the throughput results of Test-bed L, while Table 5-2 lists the average speed.

Table 5-1 20-minute throughput results of Test-bed L

<table>
<thead>
<tr>
<th></th>
<th>Left lane</th>
<th>Middle lane</th>
<th>Right lane</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>base case</td>
<td>RMC</td>
<td>base case</td>
<td>RMC</td>
</tr>
<tr>
<td>Rep. 1</td>
<td>624</td>
<td>695</td>
<td>624</td>
<td>663</td>
</tr>
<tr>
<td>Rep. 2</td>
<td>642</td>
<td>697</td>
<td>643</td>
<td>657</td>
</tr>
<tr>
<td>Rep. 3</td>
<td>617</td>
<td>680</td>
<td>636</td>
<td>654</td>
</tr>
<tr>
<td>Rep. 4</td>
<td>660</td>
<td>671</td>
<td>606</td>
<td>645</td>
</tr>
<tr>
<td>Rep. 5</td>
<td>607</td>
<td>677</td>
<td>650</td>
<td>650</td>
</tr>
<tr>
<td>Rep. 6</td>
<td>663</td>
<td>696</td>
<td>635</td>
<td>664</td>
</tr>
<tr>
<td>Rep. 7</td>
<td>621</td>
<td>693</td>
<td>620</td>
<td>662</td>
</tr>
<tr>
<td>Rep. 8</td>
<td>617</td>
<td>685</td>
<td>629</td>
<td>652</td>
</tr>
<tr>
<td>Rep. 9</td>
<td>646</td>
<td>666</td>
<td>616</td>
<td>632</td>
</tr>
<tr>
<td>Rep. 10</td>
<td>643</td>
<td>689</td>
<td>622</td>
<td>655</td>
</tr>
<tr>
<td>Average</td>
<td>634</td>
<td>684.9</td>
<td>628.1</td>
<td>653.4</td>
</tr>
<tr>
<td>% change</td>
<td>8.0%</td>
<td>4.0%</td>
<td>-1.6%</td>
<td></td>
</tr>
</tbody>
</table>

Table 5-2 20-minute average speed results of Test-bed L

<table>
<thead>
<tr>
<th></th>
<th>Left lane</th>
<th>Middle lane</th>
<th>Right lane</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>base case</td>
<td>RMC</td>
<td>base case</td>
<td>RMC</td>
</tr>
<tr>
<td>Rep. 1</td>
<td>27.4</td>
<td>32.4</td>
<td>44</td>
<td>35.9</td>
</tr>
<tr>
<td>Rep. 2</td>
<td>30.7</td>
<td>33.5</td>
<td>45.4</td>
<td>40.8</td>
</tr>
<tr>
<td>Rep. 3</td>
<td>26.7</td>
<td>30.9</td>
<td>44.6</td>
<td>40.8</td>
</tr>
<tr>
<td>Rep. 4</td>
<td>34.3</td>
<td>30.2</td>
<td>47</td>
<td>31.3</td>
</tr>
<tr>
<td>Rep. 5</td>
<td>26.4</td>
<td>30.2</td>
<td>43.3</td>
<td>31.7</td>
</tr>
<tr>
<td>Rep. 6</td>
<td>35.1</td>
<td>32.7</td>
<td>49.2</td>
<td>33.8</td>
</tr>
<tr>
<td>Rep. 7</td>
<td>28.5</td>
<td>32.3</td>
<td>45</td>
<td>34</td>
</tr>
<tr>
<td>Rep. 8</td>
<td>27</td>
<td>31.7</td>
<td>43.9</td>
<td>38.4</td>
</tr>
<tr>
<td>Rep. 9</td>
<td>30.1</td>
<td>29.7</td>
<td>47.8</td>
<td>44.4</td>
</tr>
<tr>
<td>Rep. 10</td>
<td>31</td>
<td>32.4</td>
<td>46.3</td>
<td>34</td>
</tr>
<tr>
<td>Average</td>
<td>29.7</td>
<td>31.6</td>
<td>45.7</td>
<td>36.5</td>
</tr>
<tr>
<td>% change</td>
<td>6.3%</td>
<td>-20.0%</td>
<td>-36.2%</td>
<td></td>
</tr>
</tbody>
</table>
As expected, the left lane is the one experiencing the most significant increase in throughput, by 8% from 634 veh to 684.9 veh. The middle lane also witnesses a consistent increase in throughput, by 4% from 628.1 veh to 653.4 veh. The throughput results for the right lane vary from replication to replication, and the average shows a slight decrease by -1.6% from 688.7 veh to 677.4 veh. Overall, the total merging bottleneck throughput has been improved by 3.3%, from 1950.8 veh to 2015.7 veh. However, the average speed results indicate a negative impact of RMC, and only the left lane experiences a small improvement by 6.3%, from 29.7 km/h to 31.6 km/h. The average speed of the right lane has been significantly reduced by 36.2%. In order to find the reason, the speed contour is examined (see Figure 5-3).

In Figure 5-3, the yellow rectangle indicates when and where RMC is activated. It can be clearly seen that after the activation of RMC, downstream from the bottleneck is congested, with the congestion propagating back to the merge bottleneck. When checking the new congestion location (at around 79 of the x-axis in Figure 5-3), many lane changes can be observed. This is because the Juliette Street off-ramp (at around 81 of the x-axis in Figure 5-3) is at about 400 meters downstream, and most of the diverging vehicles are starting to change their lane. As

Figure 5-3 Speed contour of Test-bed L with 20-minute RMC
the flow is increased by RMC, the density and the number of lane changes of diverging vehicles are increased, resulting in congestion. This explains why the average speed with RMC is even worse compared with the base case.

Table 5.3 20-minute throughput results of Test-bed S

<table>
<thead>
<tr>
<th></th>
<th>Left lane</th>
<th>Middle lane</th>
<th>Right lane</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>base case</td>
<td>RMC</td>
<td>base case</td>
<td>RMC</td>
</tr>
<tr>
<td>Rep. 1</td>
<td>682</td>
<td>805</td>
<td>681</td>
<td>737</td>
</tr>
<tr>
<td>Rep. 2</td>
<td>629</td>
<td>806</td>
<td>620</td>
<td>709</td>
</tr>
<tr>
<td>Rep. 3</td>
<td>645</td>
<td>789</td>
<td>637</td>
<td>738</td>
</tr>
<tr>
<td>Rep. 4</td>
<td>652</td>
<td>805</td>
<td>627</td>
<td>729</td>
</tr>
<tr>
<td>Rep. 5</td>
<td>651</td>
<td>784</td>
<td>671</td>
<td>697</td>
</tr>
<tr>
<td>Rep. 6</td>
<td>646</td>
<td>772</td>
<td>642</td>
<td>668</td>
</tr>
<tr>
<td>Rep. 7</td>
<td>638</td>
<td>792</td>
<td>619</td>
<td>701</td>
</tr>
<tr>
<td>Rep. 8</td>
<td>648</td>
<td>780</td>
<td>643</td>
<td>681</td>
</tr>
<tr>
<td>Rep. 9</td>
<td>647</td>
<td>776</td>
<td>626</td>
<td>686</td>
</tr>
<tr>
<td>Rep. 10</td>
<td>635</td>
<td>813</td>
<td>640</td>
<td>727</td>
</tr>
<tr>
<td>Average</td>
<td>647.3</td>
<td>792.2</td>
<td>640.6</td>
<td>707.3</td>
</tr>
<tr>
<td>% change</td>
<td>22.4%</td>
<td>10.4%</td>
<td>0.5%</td>
<td>10.7%</td>
</tr>
</tbody>
</table>

Table 5.4 20-minute average speed results of Test-bed S

<table>
<thead>
<tr>
<th></th>
<th>Left lane</th>
<th>Middle lane</th>
<th>Right lane</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>base case</td>
<td>RMC</td>
<td>base case</td>
<td>RMC</td>
</tr>
<tr>
<td>Rep. 1</td>
<td>52.9</td>
<td>62.2</td>
<td>63.9</td>
<td>78.4</td>
</tr>
<tr>
<td>Rep. 2</td>
<td>51.1</td>
<td>61.1</td>
<td>61.8</td>
<td>78.5</td>
</tr>
<tr>
<td>Rep. 3</td>
<td>51.9</td>
<td>66.3</td>
<td>61.7</td>
<td>78.6</td>
</tr>
<tr>
<td>Rep. 4</td>
<td>54.4</td>
<td>63.5</td>
<td>64</td>
<td>78.8</td>
</tr>
<tr>
<td>Rep. 5</td>
<td>55.9</td>
<td>62.5</td>
<td>66.1</td>
<td>78.6</td>
</tr>
<tr>
<td>Rep. 6</td>
<td>67.8</td>
<td>69</td>
<td>75.3</td>
<td>80</td>
</tr>
<tr>
<td>Rep. 7</td>
<td>53.1</td>
<td>57.7</td>
<td>63.5</td>
<td>78.9</td>
</tr>
<tr>
<td>Rep. 8</td>
<td>62</td>
<td>62.7</td>
<td>70.1</td>
<td>79.2</td>
</tr>
<tr>
<td>Rep. 9</td>
<td>54.3</td>
<td>61.6</td>
<td>63.7</td>
<td>76.6</td>
</tr>
<tr>
<td>Rep. 10</td>
<td>53.3</td>
<td>58.9</td>
<td>63.8</td>
<td>77.6</td>
</tr>
<tr>
<td>Average</td>
<td>55.7</td>
<td>62.6</td>
<td>65.4</td>
<td>78.5</td>
</tr>
<tr>
<td>% change</td>
<td>12.4%</td>
<td>20.1%</td>
<td>0.01%</td>
<td>9.9%</td>
</tr>
</tbody>
</table>
At the downstream of the Birdwood Rd. on-ramp, there is a long curve section where vehicles normally reduce their speed. As the flow suddenly increases, the density in this curve section also increases. Consequently, the speed reductions of these vehicles eventually cause the congestion. This shows how complicated the traffic flow dynamics are for such a large network.

Table 5-3 and Table 5-4 list the results for Test-bed S. The results from the simple test-bed clearly show that RMC can increase merging bottleneck throughput. Also, RMC recovers the traffic flow of improved average speed by 9.9% overall, from 67.2 km/h to 73.8 km/h. From the lane by lane results, the left lane is improved the most in throughput by 22.4% from 647.3 veh to 792.2 veh, and the middle lane is the second most affected, while the right lane is not affected much by RMC. In terms of the average speed, the middle lane witnesses the most significant improvement, by 20.1% from 65.4 km/h to 78.5 km/h. This is because there is almost no vehicle that changes lane from the left lane to the middle lane due to the huge reduction of interferences from ramp traffic.

5.5.3 Conclusions

Based on the simulation investigation, the following conclusions can be made:

- RMC can improve merging bottleneck throughput at the recovery phase from 3% to 10% in total for a three-lane motorway. The left lane experiences the most improvement and the right lane is not affected much. This indicates that the fewer number of lanes the mainline has, the more effective RM is.

- RMC can recover the mainline traffic flow at the recovery phase.

- The traffic flow dynamics are complicated for a large network. Consequently, a good RCR strategy must be able to react to the dynamics quickly.
5.6 **SUMMARY**

In this chapter, the feasibility of achieving RCR by RM is analysed and discussed. Specifically, two research questions are addressed. Firstly, the potential system benefits are identified by analysing the traffic flow features at the recovery phase. With RCR accomplished, not only mainline traffic would enjoy travel time savings, but also ramp traffic would eventually benefit from a free flow motorway network. Secondly, RMC is proposed as the basic RM operation for the RCR purpose. Analysing the impact of RMC on merging bottleneck throughput shows that RMC could accelerate the mainline queue discharge so as to achieve RCR.
Chapter 6  Zone-based Ramp Metering Strategy for Rapid Congestion Recovery

6.1 INTRODUCTION

This chapter presents the development of the strategy for rapid congestion recovery (RCR) using ramp metering (RM). The proposed strategy is named zone-based ramp metering strategy for rapid congestion recovery (ZRM-RCR). In motorway mainline queuing, each queue starts from an active bottleneck and then propagates back to upstream. The tendency of each mainline queue is strongly related to the traffic conditions from the downstream active bottleneck to the next upstream active bottleneck. Therefore, the motorway network can be divided into several zones based on those active bottlenecks. As the major objective of RCR is to clear mainline traffic queues, it is reasonable to consider recovery for every mainline queue in one zone.

In Section 6.2, zone is defined, and the framework of ZRM-RCR is developed. Section 6.3 presents the RM algorithm for recovery control. As outlined in Section 5.2.2, this research project considers two recovery scenarios: recovery from peak hour congestion and recovery from incident. The results of simulation evaluation for the two recovery scenarios are presented and discussed in Section 6.4 and Section 6.5 respectively. Section 6.6 summarises this chapter.

6.2 ZONE DEFINITION AND FRAMEWORK

This section begins by defining zone, and then introduces the framework and components in the ZRM-RCR strategy.

6.2.1 Definition of Zone

Zones are determined by the locations of active bottlenecks. Specifically, a zone is defined as a motorway section starting from one active bottleneck and finishing at the next upstream bottleneck. An example is shown in Figure 6-1.
As can be seen from Figure 6-1 there is one main bottleneck and one mainline queue in each zone. A zone can be further divided as a queuing area and a non-queuing area. The formulation and symbols of elements in a zone are defined as follows (see Figure 6-2):

- $Q_{dis}$: The merging bottleneck flow at the active bottleneck;
- $Q_{rmc}$: The merging bottleneck flow when operating RMC;
- $Q_{up}$: The upstream incoming flow at the end of the mainline queue;
- $R^i_q$: The ith on-ramp incoming flow in the queuing area;
- $R^j_n$: The jth on-ramp incoming flow not in the queuing area;
- $Q^k_q$: The ith off-ramp exiting flow in mainline queuing area.

### 6.2.2 Framework

At the concept level, only logical components are included in the framework; the flow of the control is also demonstrated. Figure 6-3 shows the framework of ZRM-RCR.
6.2.3 Zone Identification

The objective of zone identification is to identify mainline queues and to formulate each zone. The key for zone identification is to detect mainline queues and to determine the queue head and the queue tail. This project adopts a mainline queue detection algorithm based on monitoring mainline detector measurements (Bevrani, Rahman, & Chung, 2011; Lee, Chung, & Jiang, 2011; Queensland Department of Transport and Main Roads, 2008). The queue detection algorithm simply scans all detector measurements from downstream to upstream with a fixed interval (a 1-minute interval in this research project), and then determines queues by consecutive congested detectors. Details can be found in the literature (Chung, et al., 2011). This algorithm requires relatively dense detector placement along the whole motorway network; the test-bed used in this research project – the northbound Pacific Motorway, Brisbane, Australia – has an average detector spacing of 650 meters and 49 detectors in total, which makes it suitable for applying the algorithm.
With the queue detection algorithm, the queue head can be accurately determined, together with priori knowledge of any fixed bottleneck location or incident information. The exactly point of the queue tail cannot be given, but the rough location is known. Figure 6-4 shows an example of mainline queue tail detection.

As can be seen from Figure 6-4, the queue tail area is the information obtained from the detection algorithm; this is accurate enough for monitoring queue changes for the recovery phase identification. Firstly, the queue tail is moving all the time. Secondly, monitoring queue changes is based on detector measures, so the key is to select the proper detector. In the above example, DS2 is the detector used for queue monitoring.

### 6.2.4 Recovery Phase Identification

Recovery phase identification activates the recovery control. Apparently, the accuracy of recovery phase identification is critical for the ZRM-RCR. Based on the previous defined symbols for a zone, the recovery phase can be formulated as follows:

\[
Q_{up}(t) + \sum_i R_i^j(t) - \sum_k O_q^k(t) < Q_{dis}
\]

(6-1)

where “\(t\)” is the notion for time stamp and others are the same as previous.

Once inequality 6-1 is satisfied, the mainline queue length would start to reduce, indicating the recovery phase. Figure 6-5 demonstrates the moving direction of the mainline queue at the recovery phase. Also, inequality 6-1 is also used to check for a false alarm after the activation of ZRM-RCR.
In order to reduce the risk of false alarms, historical knowledge is incorporated in the module. Basically, the approximate time when the demand would start to reduce is known, so a simple way to incorporate this information is to add a time window in which the recovery phase identification is working. Another method to reduce the risk of false alarms is to put a number of consecutive intervals to confirm the recovery phase. In this research, a 2-consecutive interval is selected based on simulation calibration (see details in Section 6.4.3).

Figure 6-6 shows the flow chart of recovery phase identification. The time window is preset based on historical analysis. Once the time is on, then the process of recovery phase identification is activated. In this research project, the pre-defined time is 8:30 am for the northbound Pacific Motorway, during morning peaks.

The process logic has three steps for checking the recovery phase. The first step is to monitor mainline the incoming flow, \( Q_{up} \), at the end of the queue and to project it. The projection method is the same as in Section 4.2.3. If the flow and its projection are both reducing, the first condition is satisfied. In the second step, ramp arrival flows in the mainline queuing area are monitored and projected. Similarly, if the flow and the projection are both not increasing, the second condition is considered to be satisfied. The last step is to calculate consecutive intervals. If the two conditions are satisfied, this interval is considered to be an effective interval. Once the consecutive effective interval is over a certain threshold, it would be acknowledged as an active recovery phase and the recovery control of ZRM-RCR will be activated. The three steps are formulated as follows:

\[
\begin{cases}
\Delta Q_{up}(t) < 0 \\
\Delta Q_{up}(t+1) < 0
\end{cases}
\]  

(6-2)
\[
\begin{cases}
\Delta \sum_{i} R_{q}^{i}(t) \leq 0 \\
\Delta \sum_{i} R_{q}^{i,prj}(t + 1) \leq 0
\end{cases}
\]

\( N_{cs}(t) \geq N_{th} \) \hspace{2cm} (6-4)

where \( \Delta \) indicates the change over last interval;

Sub-script “prj” indicates projection value;

\( N_{cs} \) is the number of consecutive intervals;

\( N_{th} \) is the threshold for \( N_{cs} \).

**Figure 6-6 Flow chart of recovery phase identification**

### 6.2.5 Recovery Cancellation

Recovery cancellation module allows withdrawing ZRM-RCR for a zone or not. There are three reasons for the cancellation:

- False alarm: if the queue keeps propagating back to upstream again, this would be considered as a false alarm, and the recovery is cancelled.
6.3 TWO-PHASE CONTROL ALGORITHM

As discussed and tested in Chapter 5, restrictive metering control (RMC) is selected as the basic RM operation for recovery control. RMC can increase merging bottleneck throughput so as to accelerate the discharge of the mainline queue. Besides, RMC naturally increases ramp traffic costs, including ramp traffic travel time and queue spillover. In the recovery phase, temporary ramp traffic cost increases can trade for the accelerated recovery of the whole network, and ultimately ramp traffic can benefit from a recovered mainline traffic. Therefore, an idea for recovery is proposed. At the very beginning the system could tolerate more ramp costs than normal metering control for a short period, and then consider both mainline and ramp conditions in order to avoid excessive ramp costs. The concept of the ZRM-RCR control is stated as follows:

At the very beginning of the recovery phase, the ZRM-RCR control ignores ramp costs for a short period from a system point of view, and applies RMC to achieve an increased mainline queue discharge rate. A reactive control based on both mainline and ramp conditions would then activate to avoid unnecessary and excessive ramp costs.

The proposed ZRM-RCR control is designed to be a two-phase control: compulsory control phase followed by reactive control phase. Figure 6-7 shows the control structure.

Figure 6-7 Two-phase control structure for ZRM-RCR
6.3.1 Compulsory Control

The objective of compulsory control is to accelerate the mainline queue discharge. In a zone, there are two groups of on-ramps: those in the queuing area and those not in the queuing area. Different strategies are designed for the two groups (see Figure 6-8):

- Ramps in queuing area will run the RMC. Operating RMC at these ramps would accelerate the discharge of mainline queue.

- The other ramps will run the local RM algorithm. Before activation of ZRM-RCR, the system is running the coordinated RM. Consequently, the ramps in the non-queuing area are highly likely to be slave ramps. In this case, they might already keep a certain length of on-ramp queue. In addition, mainline demand has reduced significantly, and these ramps are away from the mainline queue. Also, the demands of these ramps are usually not very high. Accordingly, even though these ramps increase their metering rates by the local RM, the mainline queue is still likely to keep reducing. As the downstream masters (ramps in the queuing area) are forced to run RMC, it is an opportunity for the slaves to clear their own ramp queues. The benefit of clearing upstream ramp queues is that after the compulsory control phase, upstream ramps are available to become slaves with plenty of on-ramp storage space.

![Figure 6-8 Compulsory control phase](image_url)

Another important parameter for compulsory control is the length of the compulsory period. If the length is too long, it would cause unnecessary ramp costs; if it is too short, it will not have any impact. Based on simulation tests (see details in Section 6.4.3), a 3-minute period for the compulsory control phase is used for Test-bed 1 and a 5-minute period is used for Test-bed 2 in the research project.
6.3.2 Reactive Control

As the compulsory control would make ramps in the queuing area suffer from the RMC, it is not reasonable to keep it for too long time. After the compulsory control phase, the control objective is to regain control of the total ramp costs. Again, different strategies are designed for the two ramp groups (see Figure 6-9):

- Ramps in the queuing area will run the cost constrained additive increase minimal release (CC-AIMR) algorithm, a localised RM algorithm. The reason of using only local information is that even in coordinated RM, these ramps would be the master running the local RM anyway. The details of the CC-AIMR algorithm are presented below.
- The other ramps in the zone will run the coordinated RM. As these ramps already get refreshed in terms of their ramp queues, they are now able to contribute more to the system.

![Figure 6-9 Reactive control phase](image)

**CC-AIMR Algorithm**

In the CC-AIMR, the CC part is inspired from the local queue management scheme in Chapter 3, while the AIMR part is inspired by the TCP congestion control algorithm (additive increase and multiplicative decrease, AIMD algorithm). Using cost constraint will bring local ramp costs into consideration. Besides, AIMR algorithm is designed from the TCP congestion control principle, which has a quick recovery response to congestion. In the CC-AIMR algorithm, a metering rate is decided by the AIMR algorithm, while a constraint for maximum acceptable spillover is set and accumulative queue spillover is recorded. If the accumulative queue spillover is smaller than the constraint, the metering rate from AIMR will be applied; otherwise, the metering rate from the local queue management will be
applied. The flow chart is shown in Figure 6-10. Details of AIMR algorithm are introduced below.

![Flow chart of CC-AIMR algorithm](image)

**AIMR Algorithm**

As noted, the AIMR algorithm originates from the AIMD (additive increase multiplicative decrease) algorithm within TCP congestion control. The basic logic of the prototype algorithm is similar to the AIMD algorithm in TCP control. When the feedback variable still indicates a congested state in the merging area, the metering rate goes directly back to its minimum; when the feedback variable shows a recovery state, the metering rate increases additively. Therefore, this algorithm is called the additive increase minimum release (AIMR) algorithm. Compared with the AIMD algorithm, the feedback variable here is generating by the loop detector data in real time. In this research project, speed at merging area is chosen as the feedback variable, because it is the direct detector measure to check the occurrence of traffic flow breakdown. There are two operations in the AIMR algorithm:

- AI operation, which is the logic to increase metering rate additively.
- MR operation, which is the logic to set the metering rate to its minimum.

The AIMR algorithm has three steps for every calculation interval (CI, the time interval to update the metering rate, a 1-minute interval is used in this research
project) and the flowchart of the AIMR algorithm for each calculation interval is shown in Figure 6-11:

![Flow chart of AIMR algorithm](image)

**Figure 6-11 Flow chart of AIMR algorithm**

a. Get detector data.

b. Update and record state of the CI. There are three states:

   - Congested state (state C): If the speed of merge area drops below a pre-defined threshold (SD in Figure 6-11), this CI will be counted as state C.

   - Recovered state (state R): If the speed of merge area recovers to a pre-defined threshold (SA in Figure 6-11), this CI will be counted as state R.

   - Middle state (state M): If the CI state is neither state R nor state C, it will be counted as state M.

c. Determine metering rate for next CI:

   - If consecutive recovered CI satisfies a pre-defined threshold (CIA in Figure 6-11 and 2 used in this study), the metering rate for the next CI will increase by the pre-defined increment, “μ”.

Chapter 6: Zone-based Ramp Metering Strategy for Rapid Congestion Recovery
– If consecutive congested CI satisfies a pre-defined threshold (CID in Figure 6-11 and 2 used in this study), the new metering rate will be set back to the minimum metering rate.

– Otherwise, the metering rate will keep the same as in the previous CI.

6.4 EVALUATION OF PEAK HOUR CONGESTION SCENARIO

This section evaluates the performances of the proposed ZRM-RCR in the peak hour congestion scenario; that is, the morning peak congestion.

6.4.1 Simulation Settings

In this section, two test-beds are used. The first is the original northbound Pacific Motorway, the same as used in Section 4.4.1 (Test-bed 1). In Test-bed 1, the most significant bottleneck and mainline queuing are because of weaving behaviours at the Gateway Motorway interchange. The concern for the test-bed is that this might not be the best test-bed to demonstrate the effectiveness of RM, as RM is most effective for merging bottlenecks. Consequently, another test-bed, the modified northbound Pacific Motorway, the same as in Section 5.5.1 (Test-bed 2), is also used for the evaluation. Details of the modifications for Test-bed 2 can be found in Section 5.5.1.

Three test scenarios are tested in both test-beds to demonstrate the effectiveness of ZRM-RCR:

• Base case scenario assumes no RM control;

• Coordinated RM scenario (CRM) operates the coordinated RM control system as presented in Chapter 4.

• ZRM-RCR scenario operates the proposed recovery strategy after 8:30 am. Otherwise, it is the same as the CRM scenario.

The ZRM-RCR is only running at the recovery phase. Apart from the recovery phase, the ZRM-RCR is also running the CRM. Therefore, the major comparison is carried out between the CRM scenario and ZRM-RCR scenario during only the time window for activating recovery phase identification; that is, after 8:30 am.
6.4.2 Performance Indicators

A total of seven performance indicators are used in the evaluation, and their definitions are as follows:

- Total travel time (TTT): the calculation of TTT is as the same as in Section 4.3.2. Its unit is veh·h;
- Average mainline traffic delay (MTD): this indicator is chosen to demonstrate the benefit of RM, and its definition is the same as in Section 4.4.2. The unit is sec/trip;
- Average ramp traffic delay (RTD): this indicator is again the same as in Section 4.4.2. The unit is sec/veh;
- Total queue spillover time (TQST): the definition of TQST is the same as Section 4.4.2. The unit is minute.
- TTT in the recovery period (TTT-R): It is the same with TTT but only collecting for recovery period. In this research, the start point of recovery period is 8:30 am for morning peak.
- MTD in the recovery period (MTD-R): The same with MTD but only collecting for recovery period as the same in TTT-R.
- RTD in the recovery period (RTD-R): The same with RTD but only collecting for recovery period as the same in TTT-R.

6.4.3 Parameter Calibration

Two parameters (the consecutive interval for recovery phase identification and the length of compulsory control) are calibrated based on four performance indicators from simulation tests: TTT-R, MTD-R, TQST-R (similar to TQST but only collecting for recovery period) and RTD.

Consecutive Interval for Recovery Phase Identification

Three values are tested in Test-bed 1, and the results are shown in Table 6-1. The results demonstrate that the 2-minute consecutive interval gives the best performance of all four performance indicators. Accordingly, the 2-minute consecutive interval is selected in the research.
Table 6-1 Calibration results of consecutive interval for recovery phase identification

<table>
<thead>
<tr>
<th>Unit</th>
<th>2-minute</th>
<th>3-minute</th>
<th>4-minute</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT-R</td>
<td>veh/h</td>
<td>3186.7</td>
<td>3193.0</td>
</tr>
<tr>
<td>MTD-R</td>
<td>sec/trip</td>
<td>205.4</td>
<td>209.0</td>
</tr>
<tr>
<td>RTD-R</td>
<td>sec/veh</td>
<td>148.2</td>
<td>152.9</td>
</tr>
<tr>
<td>TQST-R</td>
<td>minute</td>
<td>265.5</td>
<td>272.9</td>
</tr>
</tbody>
</table>

**Compulsory Control Length**

The compulsory control length is the most important parameter for the ZRM-RCR strategy. As Test-bed 1 and Test-bed 2 have different major bottlenecks, the compulsory length is then calibrated for the two test-beds individually. Five values are tested; and the results from Test-bed 1 are illustrated in Table 6-2; Table 6-3 illustrates the results from Test-bed 2.

Table 6-2 Calibration results of compulsory control length from Test-bed 1

<table>
<thead>
<tr>
<th>Unit</th>
<th>3-minute</th>
<th>4-minute</th>
<th>5-minute</th>
<th>6-minute</th>
<th>7-minute</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT-R</td>
<td>veh/h</td>
<td>3205.0</td>
<td>3196.0</td>
<td>3192.7</td>
<td>3186.7</td>
</tr>
<tr>
<td>MTD-R</td>
<td>sec/trip</td>
<td>213.3</td>
<td>209.6</td>
<td>208.4</td>
<td>205.4</td>
</tr>
<tr>
<td>RTD-R</td>
<td>sec/veh</td>
<td>148.2</td>
<td>148.2</td>
<td>146.7</td>
<td>148.2</td>
</tr>
<tr>
<td>TQST-R</td>
<td>minute</td>
<td>246.0</td>
<td>255.5</td>
<td>266.1</td>
<td>265.5</td>
</tr>
</tbody>
</table>

It can be seen from Table 6-2 that as the compulsory control length increases, the overall system efficiency is improved (reduced TTT-R) and MTD-R reduces. This indicates that longer compulsory control accelerates the mainline recovery. In contrast, increasing the compulsory control length increases TQST-R. When comparing RTD-R results with the CRM scenario in Table 6-4, only the 7-minute scenario has a higher RTD-R than the CRM scenario, which means that ramp traffic does not benefit from quick recovery if the compulsory control phase is unnecessarily long. Overall, a 3-minute compulsory control is selected for Test-bed 1, because almost no additional spillover is caused by the ZRM-RCR.

A similar trend with Table 6-2 can be observed in Table 6-3: increased length of compulsory control improves the overall system efficiency and the mainline traffic. When checking the ramp costs (TQST-R and RTD-R), 5-minute compulsory control gives the best performance, and thereby is selected for Test-bed 2.
Table 6-3 Calibration results of compulsory control length from Test-bed 2

<table>
<thead>
<tr>
<th></th>
<th>Unit</th>
<th>3-minute</th>
<th>4-minute</th>
<th>5-minute</th>
<th>6-minute</th>
<th>7-minute</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT-R</td>
<td>veh/h</td>
<td>3436.6</td>
<td>3424.4</td>
<td>3415.3</td>
<td>3398.9</td>
<td>3390.7</td>
</tr>
<tr>
<td>MTD-R</td>
<td>sec/trip</td>
<td>673.9</td>
<td>665.2</td>
<td>660.1</td>
<td>652.4</td>
<td>649.7</td>
</tr>
<tr>
<td>RTD-R</td>
<td>sec/veh</td>
<td>229.1</td>
<td>227.4</td>
<td>223.8</td>
<td>230.6</td>
<td>233.0</td>
</tr>
<tr>
<td>TQST-R</td>
<td>minute</td>
<td>273.9</td>
<td>270.4</td>
<td>263.8</td>
<td>276.9</td>
<td>276.4</td>
</tr>
</tbody>
</table>

6.4.4 Results from Test-bed 1

In order to reduce the impact of random seed in micro-simulation, 10 replications are simulated and the results are collected. Table 6-4 summarizes the average results from 10 replications of the 7 performance indicators. In Table 6-4, there are two groups of performance indicators: the first group, including the first 4 performance indicators, is collected for the whole simulation period; the other group is collected for only the general recovery period, from 8:30 to the end of simulation.

Table 6-4 Simulation result summary of Test-bed 1

<table>
<thead>
<tr>
<th></th>
<th>Unit</th>
<th>Base case</th>
<th>CRM</th>
<th>ZRM-RCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>veh/h</td>
<td>14421.9</td>
<td>11440.2</td>
<td>11380.5</td>
</tr>
<tr>
<td>MTD</td>
<td>sec/trip</td>
<td>944.6</td>
<td>433.0</td>
<td>423.0</td>
</tr>
<tr>
<td>RTD</td>
<td>sec/veh</td>
<td>60.7</td>
<td>157.6</td>
<td>155.8</td>
</tr>
<tr>
<td>TQST</td>
<td>minute</td>
<td>228.4</td>
<td>649.4</td>
<td>651.1</td>
</tr>
<tr>
<td>TTT-R</td>
<td>veh/h</td>
<td>4406.7</td>
<td>3264.7</td>
<td>3203.3</td>
</tr>
<tr>
<td>MTD-R</td>
<td>sec/trip</td>
<td>858.8</td>
<td>242.4</td>
<td>212.3</td>
</tr>
<tr>
<td>RTD-R</td>
<td>sec/veh</td>
<td>59.4</td>
<td>153.3</td>
<td>148.0</td>
</tr>
</tbody>
</table>

As the comparison between the base case scenario and the CRM scenario of the first group has been discussed in Chapter 4, the focus here is to compare the CRM scenario and the ZRM-RCR scenario. Overall, the ZRM-RCR scenario further improves the performances over the CRM scenario, with a small cost of queue spillover time. Specifically, TTT has been further reduced by 0.5% from 11440.2 to 11380.5 veh-h, as TTT reflects the system efficiency. Considering that the ZRM-RCR is executing only a maximum 30 minutes, and the main bottleneck for Test-bed 1 is the weaving bottleneck at the Gateway Motorway interchange. This is a remarkable improvement. As expected in the saved travel time, mainline traffic experiences a 2.3% reduction in MTD, from 433 to 423 sec/trip. More interesting
results are from the RTD, in which an improvement is observed: a 1.1% decrease in RTD from 157.6 to 155.8 sec/veh, which indicates that with a quicker recovery of the whole system, ramp traffic eventually enjoys the system benefits. The TQST is almost the same, 649.4 compared with 651.1 minutes, which implies almost no additional costs for running the ZRM-RCR.

In order to highlight the benefits of ZRM-RCR, the second group of performance indicators is collected. In particular, the percentage reduction of TTT-R is 1.9% (from 3264.7 to 3203.3 veh-h). For MTD, the improvement is remarkable, with a 12.4% reduction from 242.4 to 212.3 sec/trip. This indicates that the quicker recovery provides a much better mainline traffic condition. Even for ramp traffic, the average delay in recovery period, RTD, is decreased by 3.5%. All these results verify the design principle of the ZRM-RCR: at the very beginning of the recovery phase, running RMC for a short time can assist the mainline congestion recovery; with a recovered motorway network, not only mainline traffic but also ramp traffic will benefit. Particularly considering Test-bed 1, the recovery idea is effective for a network in which the non-merging bottleneck dominates.

Table 6-5 Individual ramp queue spillover results of Test-bed 1

<table>
<thead>
<tr>
<th>Road</th>
<th>Base case</th>
<th>CRM</th>
<th>ZRM-RCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beenleigh Road</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Grandis Street</td>
<td>0</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>Murrays Road</td>
<td>0</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Centenary Road</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Loganlea Road</td>
<td>0</td>
<td>44.9</td>
<td>44.8</td>
</tr>
<tr>
<td>Service Road</td>
<td>0</td>
<td>70.0</td>
<td>75.4</td>
</tr>
<tr>
<td>Fitzgerald Avenue</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Sports Drive</td>
<td>88.8</td>
<td>244.6</td>
<td>240.4</td>
</tr>
<tr>
<td>Logan Road</td>
<td>0</td>
<td>8.4</td>
<td>7.7</td>
</tr>
<tr>
<td>Kessels Road</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Mains Road</td>
<td>0</td>
<td>13.2</td>
<td>12.8</td>
</tr>
<tr>
<td>Birdwood Road</td>
<td>0</td>
<td>51.1</td>
<td>56.6</td>
</tr>
<tr>
<td>Duke Street</td>
<td>0</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Stanley Street</td>
<td>139.6</td>
<td>205.8</td>
<td>201.5</td>
</tr>
<tr>
<td>Alice Street</td>
<td>0</td>
<td>10.6</td>
<td>11.1</td>
</tr>
<tr>
<td>Ann Street</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Table 6-5 and Table 6-6 show ramp queue spillover and ramp traffic travel time in the recovery period, for individual ramps.

Table 6-6 Individual ramp traffic travel time in the recovery period of Test-bed 1

<table>
<thead>
<tr>
<th>Ramp</th>
<th>Base case</th>
<th>CRM</th>
<th>ZRM-RCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beenleigh Road</td>
<td>41.6</td>
<td>44.5</td>
<td>43.8</td>
</tr>
<tr>
<td>Grandis Street</td>
<td>33.3</td>
<td>35.5</td>
<td>35.3</td>
</tr>
<tr>
<td>Murrays Road</td>
<td>44.8</td>
<td>45.2</td>
<td>40.6</td>
</tr>
<tr>
<td>Centenary Road</td>
<td>50.4</td>
<td>38.2</td>
<td>38.1</td>
</tr>
<tr>
<td>Loganlea Road</td>
<td>54</td>
<td>52.1</td>
<td>51.2</td>
</tr>
<tr>
<td>Service Road</td>
<td>142</td>
<td>105.1</td>
<td>111.7</td>
</tr>
<tr>
<td>Fitzgerald Avenue</td>
<td>60.6</td>
<td>32.5</td>
<td>31.8</td>
</tr>
<tr>
<td>Sports Drive</td>
<td>507.9</td>
<td>396.1</td>
<td>382.5</td>
</tr>
<tr>
<td>Logan Road</td>
<td>66</td>
<td>142.3</td>
<td>130.7</td>
</tr>
<tr>
<td>Kessels Road</td>
<td>31.9</td>
<td>57.6</td>
<td>47.6</td>
</tr>
<tr>
<td>Mains Road</td>
<td>106.8</td>
<td>197.2</td>
<td>197.3</td>
</tr>
<tr>
<td>Birdwood Road</td>
<td>86.8</td>
<td>614.6</td>
<td>614.2</td>
</tr>
<tr>
<td>Duke Street</td>
<td>49.8</td>
<td>84.5</td>
<td>81.7</td>
</tr>
<tr>
<td>Stanley Street</td>
<td>85.5</td>
<td>108.3</td>
<td>103.1</td>
</tr>
<tr>
<td>Alice Street</td>
<td>50.8</td>
<td>348.8</td>
<td>328.3</td>
</tr>
<tr>
<td>Ann Street</td>
<td>37.6</td>
<td>63.2</td>
<td>58.5</td>
</tr>
</tbody>
</table>

Individual ramp results provide inside views from which to examine the ZRM-RCR control algorithm. Based on Table 6-5, the increased queue spillover time is distributed on the Service Road on-ramp (an increase of 5.4 minutes), the Birdwood Road on-ramp (a 5.5-minute increase) and the Alice Street on-ramp (an increase of 0.5 minutes). The first two ramps are high-demand ramps in the two mainline queuing areas: the Gateway Motorway interchange bottleneck and the Birdwood Road merging bottleneck (See speed contour in Figure 4-9). In the ZRM-RCR control algorithm, these ramps are forced to operate RMC in the compulsory control phase, thereby causing more queue spillover. The situation for the Alice Street ramp is different. It is right at the downstream of the Stanley Street ramp. As the Stanley Street ramp is operating RMC at recovery phase, increased flow makes denser mainline traffic at the Alice Street ramp. Consequently, the local RM system holds more ramp traffic for mainline priority, resulting in more queue spillover.
a. The whole simulation period

b. Enlarged part of the recovery period

Figure 6-12 Motorway mainline speed contour – Test-bed 1 (CRM vs. ZRM-RCR)

When checking the individual ramp traffic travel time in Table 6-6, only slight increases in percentage as found for the Service Road ramp (6.3%) and and the Mains Road on-ramp (less than 0.1%). At both the Birdwood Road on-ramp and the Alice Street on-ramp, the average ramp traffic travel time reduces simply because the mainline queue is cleared in a short time and then ramp traffic can be quickly discharged. This indicates that the short period “pains” (maximum 5.5 minutes more
spillover at the Birdwood Road on-ramp) benefit the whole system while not increasing individual ramp travel time much.

Figure 6-12 (a) illustrates the speed contour for the whole simulation period of the CRM scenario and the ZRM-RCR scenario. As the ZRM-RCR is running for only a short time, Figure 6-12 (b) selects the period when the ZRM-RCR is activated and enlarges this part in the black rectangles in Figure 6-12 (a). As can be seen, the time duration of mainline queues in the Gateway Motorway area and the City area has been clearly reduced, which indicates a quick recovery of the mainline traffic achieved by the ZRM-RCR algorithm. Note that the speed contours in Figure 6-12 are samples; all the replications produced similar speed contour patterns.

6.4.5 Results from Test-bed 2

Test-bed 2 has only merging bottlenecks, which makes it perfect for testing RM algorithms. Therefore, another test scenario, the LRM scenario, operating the local RM control for all the 16 on-ramps along the motorway, is included in the simulation evaluation. The changed traffic patterns for Test-bed 2 are described before presenting the simulation results. Figure 6-13 displays the speed contours of the base case for both test-beds.

The two speed contours in Figure 6-13 show different traffic flow patterns. In general, the total congestion in Test-bed 2 is more severe than in Test-bed 1. Without the weaving bottleneck at the Gateway Motorway interchange, more vehicles pass the interchange and travel towards the City. Consequently, the mainline incoming traffic flows for the downstream ramps are higher than in Test-bed 1. Given the Logan Road ramp, the Mains Road ramp, the Birdwood Road ramp and the Stanley Street ramp are of high demand, their merging areas become bottlenecks causing serious mainline congestion.

Table 6-7 illustrates the result summary of all 4 test scenarios. In general, the trends are similar when comparing the LRM scenario over the base case, and the CRM scenario over the LRM scenario:

- The base case scenario results the worst overall traffic condition, the highest TTT of 15770.6 veh∙h and the highest MTD of 1169.3 sec/trip. However, the ramp costs are the smallest, including the lowest RTD of 94.3 sec/veh
and the shortest TQST of 307.6 minutes. Compared with data in Table 4-8, these figures confirm that the traffic is more congested in Test-bed 2.

![Motorway mainline speed contour of the base case scenario (Test-bed 1 vs. Test-bed 2)](image)

**Figure 6-13** Motorway mainline speed contour of the base case scenario (Test-bed 1 vs. Test-bed 2)

**Table 6-7** Simulation result summary of Test-bed 2

<table>
<thead>
<tr>
<th></th>
<th>Unit</th>
<th>Base case</th>
<th>LRM</th>
<th>CRM</th>
<th>ZRM-RCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>veh/h</td>
<td>15770.6</td>
<td>14455.4</td>
<td>13394.9</td>
<td>13233.7</td>
</tr>
<tr>
<td>MTD</td>
<td>sec/trip</td>
<td>1169.3</td>
<td>898.5</td>
<td>716.4</td>
<td>689.8</td>
</tr>
<tr>
<td>RTD</td>
<td>sec/veh</td>
<td>94.3</td>
<td>206.2</td>
<td>247.8</td>
<td>238.5</td>
</tr>
<tr>
<td>TQST</td>
<td>minute</td>
<td>307.6</td>
<td>528.8</td>
<td>717.3</td>
<td>706.6</td>
</tr>
<tr>
<td>TTT-R</td>
<td>veh/h</td>
<td>4468.9</td>
<td>3938.4</td>
<td>3576.5</td>
<td>3415.3</td>
</tr>
<tr>
<td>MTD-R</td>
<td>sec/trip</td>
<td>1107.3</td>
<td>896.6</td>
<td>730.5</td>
<td>660.1</td>
</tr>
<tr>
<td>RTD-R</td>
<td>sec/veh</td>
<td>73.6</td>
<td>224.9</td>
<td>250.9</td>
<td>223.8</td>
</tr>
</tbody>
</table>

- With the installation of the local RM system, the traffic condition has been improved, but the magnitudes of the improvements are not as significant as in Test-bed 1. In particular, TTT and MTD decrease by only 8.3% (from 15770.6 to 14455.4 veh·h) and 23.2% (from 1169.3 to 898.5 sec/trip) respectively, because the Logan Road ramp and the Mains Road ramp are
congested with an increased mainline traffic flow which cannot be handled by the local RM system.

- The CRM scenario further improves the overall traffic condition significantly. Specifically, TTT reduces remarkably over the LRM scenario by 7.3% (from 14455.4 to 13394.9 veh·h); its counterpart in Test-bed 1 is only 1%. The reduction in MTD is about 20.3% over the LRM scenario, from 898.5 to 716.4 sec/trip. These results emphasize that RM is of the highest effectiveness for merging bottlenecks.

Comparisons of mainline speed contours are provided as a visual way to demonstrate the above conclusions. Figure 6-14 compares the base case and the LRM scenario; the comparison between the LRM scenario and the CRM scenario is displayed in Figure 6-15.

![Motorway mainline speed contour – Test-bed 2 (Base case vs. LRM)](image)

The purpose of using Test-bed 2 is to highlight the effectiveness of the ZRM-RCR. Therefore, the more important comparison is between the CRM scenario and the ZRM-RCR scenario. Surprisingly, all the seven performance indicators are improved by the ZRM-RCR strategy over the CRM scenario. Even for the TQST, a slight reduction is observed by 1.5% from 717.3 to 706.6 minutes, because RMC can increase the maximum mainline queue discharge rate if the main bottlenecks are merging bottleneck (see Section 5.4 and Section 5.5). This extra capacity improves the system performance with even fewer costs.
As expected, the improvements in other performance indicators are higher than in Test-bed 1, which indicates that ZRM-RCR can achieve the highest effectiveness for congestion caused by merging bottlenecks. Table 6-8 displays the performance indicator comparison between the CRM scenario and the ZRM-RCR scenario.

Table 6-8 Comparison of the performance indicators (CRM vs. ZRM-RCR)

<table>
<thead>
<tr>
<th>Unit</th>
<th>CRM</th>
<th>ZRM-RCR</th>
<th>Reduction</th>
<th>Percentage reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>veh/h</td>
<td>13394.9</td>
<td>13233.7</td>
<td>161.2</td>
</tr>
<tr>
<td>MTD</td>
<td>sec/trip</td>
<td>716.4</td>
<td>689.8</td>
<td>26.6</td>
</tr>
<tr>
<td>RTD</td>
<td>sec/veh</td>
<td>247.8</td>
<td>238.5</td>
<td>9.3</td>
</tr>
<tr>
<td>TQST</td>
<td>minute</td>
<td>717.3</td>
<td>706.6</td>
<td>10.7</td>
</tr>
<tr>
<td>TTT-R</td>
<td>veh/h</td>
<td>3576.5</td>
<td>3415.3</td>
<td>161.2</td>
</tr>
<tr>
<td>MTD-R</td>
<td>sec/trip</td>
<td>730.5</td>
<td>660.1</td>
<td>70.4</td>
</tr>
<tr>
<td>RTD-R</td>
<td>sec/veh</td>
<td>250.9</td>
<td>223.8</td>
<td>27.1</td>
</tr>
</tbody>
</table>

Based on the aggregated performance indicators in Table 6-8, it seems as if the improvement does not bring any extra cost. RMC does provide some extra capacity for mainline queue discharge, but is it really a “free lunch”? In order to answer this question, individual ramp results are examined for the CRM scenario and the ZRM-RCR scenario, particularly individual ramp queue spillover time in Table 6-9 and individual average ramp traffic travel time in Table 6-10.
Table 6-9 Individual ramp queue spillover time comparison (CRM vs. ZRM-RCR)

<table>
<thead>
<tr>
<th></th>
<th>CRM</th>
<th>ZRM-RCR</th>
<th>Difference (CRM – ZRM-RCR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beenleigh Road</td>
<td>0.3</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>Grandis Street</td>
<td>1.4</td>
<td>1.3</td>
<td>0.1</td>
</tr>
<tr>
<td>Murrays Road</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Centenanny Road</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Loganlea Road</td>
<td>15.7</td>
<td>15.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Service Road</td>
<td>30.4</td>
<td>30</td>
<td>0.4</td>
</tr>
<tr>
<td>Fitzgerald Avenue</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sports Drive</td>
<td>152.2</td>
<td>146.2</td>
<td>6</td>
</tr>
<tr>
<td>Logan Road</td>
<td>134.8</td>
<td>142.4</td>
<td>-7.6</td>
</tr>
<tr>
<td>Kessels Road</td>
<td>7.8</td>
<td>7.8</td>
<td>0</td>
</tr>
<tr>
<td>Mains Road</td>
<td>90.6</td>
<td>88.5</td>
<td>2.1</td>
</tr>
<tr>
<td>Birdwood Road</td>
<td>59.4</td>
<td>73.8</td>
<td>-14.4</td>
</tr>
<tr>
<td>Duke Street</td>
<td>1.5</td>
<td>0</td>
<td>1.5</td>
</tr>
<tr>
<td>Stanley Street</td>
<td>210.5</td>
<td>189.9</td>
<td>20.6</td>
</tr>
<tr>
<td>Alice Street</td>
<td>10</td>
<td>10.8</td>
<td>-0.8</td>
</tr>
<tr>
<td>Ann Street</td>
<td>2.7</td>
<td>0.1</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Three minus figures, which indicate more queue spillover in the ZRM-RCR scenario over the CRM scenario, are observed in Table 6-9: -7.6 minutes at the Logan Road ramp, -14.4 minutes at the Birdwood ramp and -0.8 at the Alice Street ramp. The first two ramps are related to the mainline queue starting from the Birdwood merging bottleneck, as displayed in Figure 6-13. They are high-demand ramps in the queuing area even at recovery phase, so the RMC leads to more queue spillover. Especially for the Birdwood Road ramp, the upstream mainline queue is too long to be discharged with only 5-minute RMC operation. Consequently, ramp traffic here cannot be quickly discharged due to the ongoing congestion of the mainline at the Birdwood merging area. For the Alice Street ramp, the situation is similar to Test-bed 1. Interestingly, the Stanley Street ramp witnesses a reduction in queue spillover time, from 210.5 to 189.9 minutes. As one main bottleneck, it is supposed to pay for the RMC operation. The situation is that its own demand reduces at the recovery phase, so RMC does not increase ramp queue much. Then, the recovered mainline makes it possible to quickly discharge all the previous queuing vehicles. Consequently, less queue spillover is observed.
Table 6-10 Individual average ramp traffic travel time at the recovery period comparison (CRM vs. ZRM-RCR)

<table>
<thead>
<tr>
<th></th>
<th>CRM</th>
<th>ZRM-RCR</th>
<th>Difference (CRM – ZRM-RCR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beenleigh Road</td>
<td>61.5</td>
<td>59.4</td>
<td>2.1</td>
</tr>
<tr>
<td>Grandis Street</td>
<td>37.7</td>
<td>38.8</td>
<td>-1.1</td>
</tr>
<tr>
<td>Murrays Road</td>
<td>42.8</td>
<td>40.7</td>
<td>2.1</td>
</tr>
<tr>
<td>Centennary Road</td>
<td>41.2</td>
<td>40.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Loganlea Road</td>
<td>59.2</td>
<td>61.9</td>
<td>-2.7</td>
</tr>
<tr>
<td>Service Road</td>
<td>78.2</td>
<td>76</td>
<td>2.2</td>
</tr>
<tr>
<td>Fitzgerald Avenue</td>
<td>18.8</td>
<td>19.2</td>
<td>-0.4</td>
</tr>
<tr>
<td>Sports Drive</td>
<td>297.6</td>
<td>268.2</td>
<td>29.4</td>
</tr>
<tr>
<td>Logan Road</td>
<td>458.9</td>
<td>468.9</td>
<td>-10</td>
</tr>
<tr>
<td>Kessels Road</td>
<td>205.6</td>
<td>197.8</td>
<td>7.8</td>
</tr>
<tr>
<td>Mains Road</td>
<td>1008.7</td>
<td>802.8</td>
<td>205.9</td>
</tr>
<tr>
<td>Birdwood Road</td>
<td>871.9</td>
<td>883.9</td>
<td>-12</td>
</tr>
<tr>
<td>Duke Street</td>
<td>120.4</td>
<td>103.2</td>
<td>17.2</td>
</tr>
<tr>
<td>Stanley Street</td>
<td>121.5</td>
<td>90.5</td>
<td>31</td>
</tr>
<tr>
<td>Alice Street</td>
<td>459.8</td>
<td>453.5</td>
<td>6.3</td>
</tr>
<tr>
<td>Ann Street</td>
<td>110.1</td>
<td>86.6</td>
<td>23.5</td>
</tr>
</tbody>
</table>

When checking the individual ramp traffic travel time in Table 6-10, only slight increases are observed, even for the Logan Road ramp (2.2% increase) and the Birdwood Road ramp (1.4% increase). Eventually, almost all the ramp traffic benefits from a quicker system recovery.

Figure 6-16 compares the mainline speed contour between the CRM scenario and the ZRM-RCR scenario to demonstrate the effectiveness of ZRM-RCR, including the whole simulation period and the enlarged recovery phase. In Figure 6-16(b), three zones, generated dynamically by the algorithm, are marked by black dashed lines. With the ZRM-RCR control algorithm, mainline conditions in the three zones recover quicker than in the CRM scenario.

6.4.6 Macroscopic Fundamental Diagram Analysis for Test-bed 2

Test-bed 2 is the network in which the major bottlenecks are due to merge. In this condition, activating RMC at the recovery phase is expected to obtain extra capacity at the merging bottleneck. The purpose of this analysis is to demonstrate this phenomenon at a system level.
The macroscopic fundamental diagram (MFD) was first proposed by Daganzo (2005, 2007), who recognised that traffic in a large network can be modelled dynamically at an aggregated level. Geroliminis and Daganzo (2007, 2008) verified the existence of MFD using Yokohama data, and Geroliminis and Sun (2011) analysed the MFD for motorway networks. Although real data analysis showed that MFD in motorway networks is of high scatter and exhibits hysteresis phenomena, MFD is able to evaluate motorway traffic conditions at a system level.

Figure 6-16 Motorway mainline speed contour – Test-bed 2 (CRM vs. ZRM-RCR)
In this analysis, the flow rate and density are aggregated for every 5 minutes, and the MFD is defined as the weighted average flow rate against the average density of the network (only for mainline traffic), based on the simulation data:

\[
\bar{q}^{w} = \frac{\sum_{i} q_{i} \cdot l_{i}}{\sum_{i} l_{i}}
\]

(6-5)

\[
\bar{k} = \frac{\sum_{i} k_{i}}{\sum_{i} 1}
\]

(6-6)

where subscript “\(i\)” represents the section index;

\(\bar{q}^{w}\) is the weighted average flow rate of the network;

\(q\) is section flow rate;

\(l\) is section length;

\(\bar{k}\) is the average density of the network;

\(k\) is section density.

Figure 6-17 MFD demonstration - Test-bed 2 (Base case and LRM)
Figure 6-17 displays the MFDs for the base case and the LRM scenario. In the MFD, the arrows indicate the time sequences of the dots. Blue dots are before 9:00 am (before the recovery), and red dots are approximately in the recovery period. It can be seen that the maximum flow rate in the LRM scenario (over 5000 veh⋅h) is slightly higher than in the base case (just around 5000 veh/h). Once the congestion has been built, the flow rate stays almost stable in both scenarios. This is because once congestion happens the network flow rate is dominated by the bottleneck throughput. With the brown broken line, it is clear that the maximum density in the LRM scenario is smaller, which means the maximum mainline queue length is shorter (this can be confirmed by cross-checking speed contours in Figure 6-14).

Figure 6-18 MFD comparison – Test-bed 2 (CRM vs. ZRM-RCR)

The MFDs of the CRM scenario and the ZRM-RCR scenario are compared in Figure 6-18. Two important phenomena can be observed from the MFDs. Firstly, the flow rate during the recovery period in the ZRM-RCR scenario is higher than its counterpart in the CRM scenario (marked in brown dashed rectangle). This is an
evidence that RMC can improve merging bottleneck throughput. Secondly, the dots in the purple circle represent the network condition at 10:00 am. As can be seen, the density of the ZRM-RCR scenario is lower than in the CRM scenario, which implies that the system has recovered more quickly by the ZRM-RCR strategy.

In order to better illustrate the effects of ZRM-RCR, only dots after 8:30 am are used for plotting the MFDs. The MFDs of the two scenarios are then put in the same coordinate system, as shown in Figure 6-19. This clearly shows how the ZRM-RCR works. With RMC operating at the compulsory control phase, the system flow rate has been increased (see the black ellipse in Figure 6-19). Although the system flow rate slight reduces in the reactive control phase, it is still higher than in the CRM scenario. This quickly reduces the system density, a sigh of system recovery. Then, everything goes much smoother.

![Figure 6-19 MFDs during the recovery period (CRM vs. ZRM-RCR)](image)

In summary, the MFD analysis confirms the effects of RMC operation for merging bottlenecks, illustrating the mechanism of the ZRM-RCR.
6.5 EVALUATION OF INCIDENT SCENARIO

6.5.1 Incident Settings

This section evaluates the performances of the proposed ZRM-RCR in the incident scenario. This scenario artificially creates an incident for a part of the simulation period. The artificial incident can be generated in peak hours or in non-peak periods. For peak hours, the network has already been affected by serious congestion caused by high demand. To insert an incident would just increase the severity of congestion. As congestion in peak hours has been discussed in previous section, this section models a non-peak period traffic condition from noon to 3:00 pm. The first step is to determine the severity for the incident. During a non-peak period, a light incident cannot affect traffic condition significantly, so a severe incident is selected.

This section creates an incident that blocks the two mainline lanes out of the three at approximately 1200 meters downstream from the Sports Drive ramp of the northbound Pacific Motorway (Test-bed 1). The incident starts from 1:00 pm and ends at 1:20 pm.

6.5.2 Recovery Phase Identification for Incident Scenario

The traffic dynamics at the recovery phase of the incident scenario are different from the congestion scenario. Consequently, the identification of the recovery phase needs to be adjusted.

![Figure 6-20 Queue propagation and dissipation example for incident scenario (Lee, et al., 2011)](image-url)
When an incident is cleared, the capacity at the incident location suddenly increases dramatically. Particularly for this incident scenario, two blocked lanes suddenly become usable out of a three-lane motorway, which means the capacity is three time that available during the incident. The capacity increase will generate a backward shockwave of queue dissipation. Figure 6-20 illustrates an example of queue propagation and dissipation caused by an incident. At time A, an incident happens, causing queue propagation back to upstream; at time B, the incident is cleared and a backward shockwave is generated by the sudden increased capacity; once the queue dissipation shockwave catches up with the queue propagation shockwave at time C, the whole mainline queue is cleared.

Based on the above analysis, the recovery phase for each ramp in the queuing area is the time when the queue dissipation shockwave arrives at its merging area. In this research, the incident clear time is assumed to be available. Figure 6-21 shows the recovery phase identification for the incident scenario. After the incident is cleared, each ramp in queuing area will monitor the change of the flow rate at the downstream merging area. Once there is a significant increase in the flow rate, this is considered as the recovery phase for this ramp.

![Flow chart of recovery phase identification for incident scenario](image)

Figure 6-21 Flow chart of recovery phase identification for incident scenario

The two-phase control algorithm is adjusted according to the changed recovery phase identification. Once the recovery phase for one ramp is detected, the ramp activates the two-phase control algorithm. Another difference is that upstream ramps in the non-queuing area are always running the CRM. Figure 6-22 demonstrate the process.

### 6.5.3 Simulation Settings

The simulation settings are similar to Section 6.4.1. Three test scenarios are tested, namely the base case, the CRM scenario and the ZRM-RCR scenario. Only 4 performance indicators are used: TTT, MTD, RTD and TQST (see Section 6.4.2).
For MTD, this section collects travel delays of only the motorway sections from the incident location to the network entrance, because the downstream of the incident location is always in a free flow condition.

Figure 6-22 ZRM-RCR for the incident scenario

### 6.5.4 Result Analysis

Table 6-11 summarizes the simulation results for the three test scenarios.

Table 6-11 Simulation result summary of the incident scenario

<table>
<thead>
<tr>
<th>Unit</th>
<th>Base case</th>
<th>CRM</th>
<th>ZRM-RCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>veh/h</td>
<td>4020.7</td>
<td>3551.2</td>
</tr>
<tr>
<td>MTD</td>
<td>sec/trip</td>
<td>465.3</td>
<td>313.9</td>
</tr>
<tr>
<td>RTD</td>
<td>sec/veh</td>
<td>26.1</td>
<td>85.5</td>
</tr>
<tr>
<td>TQST</td>
<td>minute</td>
<td>0</td>
<td>61.7</td>
</tr>
</tbody>
</table>
With the CRM running, the overall traffic condition improves significantly over the base case. Particularly, TTT reduces by 11.7% from 4020.7 to 3551.2 veh·h, while MTD witnesses a huge reduction of 32.5% from 465.3 to 313.9 sec/trip. On the other hand, the CRM introduces much higher RTD and TQST: increments of 59.4 sec/veh and 61.7 minutes respectively. Note that the incident information is blind to the CRM system; therefore, the CRM might not achieve the best results. However, the CRM still reduces mainline congestion dramatically, as illustrated in the mainline speed contour comparison presented in Figure 6-23.

The modified ZRM-RCR strategy further improves the overall traffic condition slightly (a 2.5% decrease in TTT from 3551.2 to 3461.8 veh·h), but reduces MTD significantly by 10.4% from 313.9 to 281.2 sec/trip. The improvement is obtained by a 30.2-minute increase in TQST from 61.7 to 91.9 minutes and an increase of 11.7% in the RTD. From the data, the ratio of effectiveness over cost is not as good as in the congestion scenario. Firstly, this is partly because the recovery phase identification for incident does not consider ramp demands. Secondly, it is still acceptable because the absolute amount of queue spillover is small. There are 6 on-ramps involved, so the average increase is only 5 minute/ramp. In addition, non-peak period is supposed to provide a free flow mainline; therefore, the mainline traffic is of high priority.
According to the mainline speed contour comparison displayed in Figure 6-24, the ZRM-RCR does provide a much better recovery.

Figure 6-24 Motorway mainline speed contour – Incident scenario (CRM vs. ZRM-RCR)
6.6 SUMMARY

In this chapter, a strategy for RCR using RM, namely the zone-based ramp metering strategy for rapid congestion recovery (ZRM-RCR), is developed. The strategy consists of four logic components. Especially for the recovery RM control, a two-phase control algorithm is designed to rapidly recover the mainline congestion and to control the ramp costs at the same time. The strategy is then simulated for evaluation in two recovery scenarios: peak hour congestion and incident. Especially for the congestion scenario, two test-beds are developed to highlight the effectiveness of the proposed strategy, and MFD is also used to demonstrate its impact at the system level. The following conclusions are drawn based on evaluation results:

- The proposed strategy provides a better mainline congestion recovery at the recovery phase; it improves mainstream traffic condition significantly and ultimately benefits the ramp traffic.
- The maximum benefits can be achieved if merging bottlenecks are the major cause of congestion. In this condition, extra capacity is obtained by operating RMC at the compulsory control phase, thereby obtaining extra benefits.
- The MFD analysis confirms that RMC can improve merging bottleneck throughput by the observation of the increased network flow rate in the MFD.
- The on-ramps at the major bottlenecks satisfy for quicker recovery.
- With proper modification in the recovery phase identification for incident, the proposed control algorithm can assist recovery after the incident clearance.
Chapter 7 Evaluation using a Modified Cell Transmission Model

7.1 INTRODUCTION

In Chapter 5, the recovery concept was analysed; in Chapter 6 a ramp metering (RM) strategy for rapid congestion recovery (RCR) based on zone (ZRM-RCR) was developed and tested. One concern is that all these analyses and evaluation results rely on the micro-simulation model in the AIMSUN platform. Although the model has been calibrated with field data (Chung, et al., 2011) and AIMSUN is a widely used commercial micro-simulation platform, it is important to crosscheck the concept of the proposed strategy in a different model. This would further confirm the concept and its effectiveness for congestion recovery.

Dynamic traffic flow models are a key component for developing and testing applications of intelligent transportation systems (ITS), especially for real-time traffic control systems, such as RM. This is because real-time traffic control systems usually require a large investment in infrastructure when tested or applied in the field. Consequently, evaluation on a simulation model is critical before any field tests. To model complex motorway traffic, there are two types of models, namely the microscopic (e.g. car-following) model and the macroscopic (e.g. hydrodynamic-based) model. A microscopic model, such as AIMSUN, describes the vehicle (car) following behaviour and the lane changing behaviour of every individual vehicle in the traffic. Evaluation through a microscopic model gives a way of examining the impacts on individual vehicles. A macroscopic model looks at the mathematical relationship among traffic flow characteristics, such as density, flow rate, mean speed of a traffic stream. Evaluation using a macroscopic model provides an analytical understanding from a traffic flow point of view. Therefore, a comprehensive evaluation should include both a microscopic model and a macroscopic model. Therefore, the purpose of this chapter is to test the recovery concept and evaluate the proposed RM strategy using a macroscopic model.

The cell transmission model (CTM) was first proposed by Daganzo (1994, 1995). Not only is CTM easy to formulate for and apply to a motorway network, it is
also able to reproduce major aspects of traffic’s evolution over time and space, including transient phenomena such as the building, propagation, and dissipation of queues (Daganzo, 1994). These two features make CTM a widely used macroscopic modelling tool that has been successfully applied in traffic state estimation (A. H. Ghods, Fu, & Rahimi-Kian, 2010; H.-j. Zhang, Yang, & Zhang, 2005) and in active traffic control analyses (Esmaeili, Sadeghi, & Fesharaki, 2014; Jiang, Chung, & Lee, 2012; Levinson & Zhang, 2006; Shang, Huang, & Gao, 2007; Srivastava & Geroliminis, 2013; L. Zhang, 2007). Therefore, this research project selects CTM as the macroscopic model.

Field studies showed that the merging capacity is affected significantly by ramp traffic, as analysed in Chapter 5. However, to simplify the model, the standard CTM (Daganzo, 1994, 1995) considers a fixed capacity for the merging cell, as the impact of ramp traffic on merging capacity is difficult to quantify. To better evaluate RM in CTM, this chapter attempts to introduce variable merging capacity that is related to metered ramp flow rate. A hybrid modelling framework is proposed for relating merging capacity and ramp flow rate to apply variable merging capacity at merge cell, and the CTM with variable merging capacity is the Modified CTM (M-CTM). The M-CTM is then used for evaluating the ZRM-RCR strategy in Section 7.4; evaluation results from the M-CTM are compared with those from the micro-simulation. Conclusions are summarised in Section 7.5.

7.2 MERGE CELL IN CTM AND ITS WEAKNESS

CTM is a “time-scan” strategy. In each simulation interval, a cell updates its own state with its own sending flow and the receiving flow from the upstream cell. In other words, the cell has to determine the sending rate with its downstream cell and the receiving rate with its upstream cell. An on-ramp can be modelled by CTM as shown in Figure 7-1, with 4 cells: the upstream mainline cell, the on-ramp cell, the merge cell and the downstream mainline cell.

The sending rate of the merge cell needs to be determined separately because it has two different sources. The receiving flow already has merge models, proposed by Daganzo in the literature (Daganzo, 1995), that consider the receiving function of the merge cell from two sources (mainline upstream and ramp traffic).
In the literature (Daganzo, 1995), however, the sending function of the merge cell, which describes the relationship between the merge cell and its downstream cell, is not well discussed and modelled. For an on-ramp with a one-lane mainline in reality, its maximum throughput from the merge point to its downstream (representing the capacity) should not be a fixed value; it should be a variable that is related to the ratio of the mainline traffic over the ramp traffic. Mainline traffic is of high density during the peak hours, and every merging attempt from the ramp traffic can be considered as a disturbance to the mainline traffic. Therefore, the merge capacity should be negatively correlated to the ramp flow rate.

No modification of the sending function has been found in the literature review for this thesis. Therefore, the sending rate of merge cell would be decided by the standard sending function in CTM as follows:

\[
y_{merge}(t) = \min(n_{merge}(t), Q_{merge}(t), \delta[N_{dn} - n_{dn}(t)])
\] (7-1)

where “\(t\)” is time step;

“\(y\)” is the sending flow to downstream cell;

“\(n\)” is the number of vehicles in a cell, that is calculated from Equation 7-2;

“\(Q\)” is the maximum number of vehicles that can flow into downstream cell when clock advances from \(t\) to \(t+1\), which is decided by a triangle fundamental diagram for each cell (see Figure 7-2);

“\(N\)” is the maximum number of vehicles that can be present in a cell, and it is a product of cell length and the jam density in Figure 7-2;
Subscript “merge” represents merge cell in Figure 7-1, while subscript “dn” represents downstream cell.

In Figure 7-2, “v”, “w” and “$k_{jam}$” are constants denoting, respectively, the free-flow speed, the backward wave speed and the jam density.

\[ n_{merge}(t) = n_{merge}(t-1) + \gamma_{up}(t-1) + \gamma_{r}(t-1) - y_{merge}(t-1) \]  

(7-2)

where subscript “up” represents upstream cell; subscript “r” represents ramp cell.

The merge cell state is updated by the following equation:

The physical means of the three terms in Equation 7-1, which are used to determine the sending rate for merge cell, are summarised as follows:

- Term 1, “$n_{merge}(t)$”, indicates demand for the sending flow;
- Term 2, “$Q_{merge}(t)$”, represents the merge capacity. Given the triangular fundamental diagram, it is usually set as a fixed value;
- Term 3, “$N_{dn} - n_{dn}(t)$”, reflects the available space at the downstream cell.

Given the high demands of both mainline and ramp during peak hours, Term 1 is highly likely to be a larger value over Term 2. Assuming that the merge bottleneck is the only active bottleneck and that there is no downstream congestion propagated back, the downstream cell is always in a free flow condition. As a result, Term 3 is always higher than Term 2. Therefore, the sending rate of merge cell is always equal to the merging capacity during peak hours, which is a fixed value in the standard
CTM. However, using a fixed out-flow rate during peak hours is too simple for modelling the merge cell during peak hours, especially when the on-ramp is metered. It cannot properly represent the impact of RM on the merging area. Firstly, the merging capacity is strongly related to the ramp traffic rate in a dense condition, at least for the leftmost lane. The higher the ramp traffic rate is, the more interferences the mainline traffic will receive. Consequently, a lower merging capacity is expected. Secondly, a metered ramp runs the one-car-per-green principle in Australia. With this signal setting, the impact of each merge from the ramp on the mainline traffic is supposed to be smaller than a platoon entering. This difference at the micro level would have effects on merging capacity at the macro level, and a fixed merging capacity cannot represent the impact. From the above, it is important to consider the impact of merging behaviours on the merging bottleneck so that the analytical model can better describe RM’s impact.

7.3 VARIABLE MERGING CAPACITY RELATING TO METERING RATE

The merging capacity is affected by many factors, such as weather conditions and driving behaviours. Among all these factors, the disturbance caused by ramp merging traffic is the dominant factor. This is why the merging area can easily become a bottleneck in the motorway. Therefore, this section takes the ramp merging traffic rate as the major factor affecting merging bottleneck, and proposes a hybrid modelling framework to quantify the relationship between the merging capacity and the metering rate (assuming the on-ramp is metered). As merging vehicles can be treated as disturbances to the mainstream traffic, this section begins by presenting a theory, the moving bottleneck theory, which can be used for quantifying the impact.

7.3.1 Moving Bottleneck Theory

The moving bottleneck theory has been proposed initially for modelling kinematic wave traffic streams containing slow vehicles, such as slow trucks (Daganzo & Laval, 2005a, 2005b). All these slow vehicles can be modelled discretely as moving boundaries that affect the traffic stream. Laval and Daganzo then applied the moving bottleneck theory for modelling lane-changing in traffic stream (J. A. Laval & Daganzo, 2006). In their work, lane-changing vehicles are treated as discrete particles with bounded accelerations (moving bottlenecks) in their new lane. The trajectories of these discrete particles can be determined endogenously.
using the constrained-motion (CM) model of vehicle dynamics. Laval et al. have used this concept to model merges, as merge vehicles are forced lane-changing vehicles (J. Laval, Cassidy, & Daganzo, 2007). The acceleration of discrete particles in the CM model is given as follows (J. Laval, et al., 2007):

$$a_{max} = 3.4 \cdot \left(1 - \frac{v}{v_{max}}\right) \text{ m/s}^2$$  \hspace{1cm} (7-3)

where “$a_{max}$” is the acceleration;

“$v$” is current speed;

“$v_{max}$” is the maximum speed (free-flow speed in this section).

This section adopts the concept of modelling lane-changing vehicles as moving bottleneck with bounded acceleration in the mainline traffic streams. The next step is to build a model to simulate the mainline traffic stream with moving bottleneck.

### 7.3.2 Linear Car-Flowing Model

One simplest car-following model, CF(L) (Herman, Montroll, Potts, & Rothery, 1959), relates speed and spacing by a linear rule, with a time lag. Daganzo (Daganzo, 2006) has proved that the vehicle trajectories predicted by the CF(L) are the same as those produced with the kinematic wave model with a Triangular fundamental diagram (KW(T)). As previously noted, the KW(T) model simulates traffic stream except moving bottlenecks. Therefore, the CF(L) can also be used to simulate the vehicles’ trajectories in mainline traffic streams, except moving bottlenecks. CF(L) is given as follows:

$$v^i_{n+1} = v(s^n_i) = \min\left(\frac{s^n_i - \delta}{\tau}, v_f\right), s^n_i \geq \delta$$ \hspace{1cm} (7-4)

$$x^i_{n+1} = \min(x^n_i + \tau v_f, x^{n-1}_i - \delta)$$ \hspace{1cm} (7-5)

where “$n$” is vehicle index from downstream to upstream;

“$i$” is time step;

“$s$” is the space from a follower to its leader;

“$\delta$” is the minimum possible spacing;

“$\tau$” is a sensitivity coefficient with units of time that is assumed to be sampling interval, and it is given by Equation 7-6;
“x” is vehicle location.

\[ \delta_0 = \delta + \tau v_f \]  

(7-6)

where “\( \delta_0 \)” is the flow-maximizing spacing (spacing at critical density).

### 7.3.3 Hybrid Modelling Framework

The hybrid modelling framework first assumes a free-flow speed traffic stream in which all the vehicles hold the flow-maximizing spacing from each other. Once there is a merging vehicle, a discrete particle (moving bottleneck) is generated at the merge area with initial speed “\( \theta v \)” and bounded acceleration given by Equation 7-3. Note that, “\( \theta \)”, the only parameter introduced in this model in range of \([0, 1)\), is called a disturbance coefficient. The ramp traffic is assumed to be metered under the one-car-per-green principle; that is, only one merge vehicle every green time, with a fixed frequency that is determined by the metering rate. By doing this, a disturbance is inserted into the traffic stream, and the trajectories calculated by the model are affected by the disturbances. Then, the vehicle count obtained from trajectories at the downstream of merging area is considered to be the merging capacity under certain metering rate. With this process, samples can be obtained by simulating merging capacity under different metering rates. Finally, the function of merging capacity by metering rate can be generated by regression from sample data. The flow chart is shown in Figure 7-3.

![Flow chart of generating the relationship between merging capacity and metering rate](image-url)
The hybrid modelling framework can be summarized from the microscopic level to the macroscopic level. The framework considers the most significant source of interferences for the mainstream traffic – merging vehicles from ramp to mainline, and quantifies the impact by a simple microscopic model. A numeral relationship between merging capacity and ramp flow rate, which contains the impact at the microscopic level, is extracted from the micro-simulation results. With the numeral relationship, the framework can provide variable merging capacity for merge cells in CTM at a macroscopic level.

Other mainline lanes, except the leftmost lane, will witness some lane changes because the drivers do not want to be affected by the merging vehicles. In this section, the impacts of these lane changes are not considered for simplicity. Therefore, this section focuses on the merging capacity of the leftmost lane.

7.3.4 Model Settings and Results

The critical density for maximizing flow rate, \(k_c\), is set as 20 veh/km. Assuming the average vehicle length is 4 meters and the minimum spacing between vehicles, \(\delta\), is 1 meter; the jam density, \(k_{jam}\), is 200 veh/km. Accordingly, the flow-maximizing spacing, \(\delta_0\), is 46 meters. The free flow speed, \(v_f\), is 100 km/h. In the simulation, the maximum speed, \(v_{max}\), is assumed to be equal to the free flow speed of 100 km/h. The sensitivity coefficient and the sampling interval for the CF(L), \(\tau\), is calculated to be 1.618705 second by Equation 7-6. The length of the merging lane is 300 meters. All the parameters are summarized in Table 7-1.

Table 7-1 Model parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k_c)</td>
<td>20 veh/km</td>
<td>Critical density</td>
</tr>
<tr>
<td>(k_{jam})</td>
<td>200 veh/km</td>
<td>Jam density</td>
</tr>
<tr>
<td>(\delta)</td>
<td>1 meter</td>
<td>Minimum possible spacing</td>
</tr>
<tr>
<td>(\delta_0)</td>
<td>46 meter</td>
<td>Flow-maximizing spacing</td>
</tr>
<tr>
<td>(v_f = v_{max})</td>
<td>100 km/h</td>
<td>Free flow (Maximum) speed</td>
</tr>
<tr>
<td>(\tau)</td>
<td>1.618705 sec</td>
<td>Sensitivity coefficient of CF(L)</td>
</tr>
<tr>
<td>(l_{merging-lane})</td>
<td>300 meter</td>
<td>Length of merging area</td>
</tr>
</tbody>
</table>

The disturbance coefficient, \(\theta\), represents the magnitude of the merging disturbance on the mainline traffic stream. As there is no field data for its calibration,
it is assumed to be from 0.4 to 0.8, with 0.1 as increment for sensitivity analysis. Another issue regarding the disturbance coefficient is whether it is a fixed value. In reality, the value is related to the traffic stream state and it should change from time to time and from one vehicle to another. However, it is difficult to apply a dynamic \( \theta \). In addition, this study assumes the same merging behaviour for each vehicle; note that the signal is under the one-car-per-green principle. With these two conditions, a uniform the impact of disturbance is assumed in this study. Accordingly, using a fixed disturbance coefficient is a reasonable simplification for the study.

Table 7-2 Simulation results summary

<table>
<thead>
<tr>
<th>Metering rate (veh/h)</th>
<th>Merging capacity (veh/h) θ=0.4</th>
<th>θ=0.5</th>
<th>θ=0.6</th>
<th>θ=0.7</th>
<th>θ=0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>171</td>
<td>1707</td>
<td>1740</td>
<td>1775</td>
<td>1827</td>
<td>1868</td>
</tr>
<tr>
<td>185</td>
<td>1692</td>
<td>1725</td>
<td>1758</td>
<td>1812</td>
<td>1857</td>
</tr>
<tr>
<td>202</td>
<td>1678</td>
<td>1710</td>
<td>1744</td>
<td>1798</td>
<td>1847</td>
</tr>
<tr>
<td>222</td>
<td>1644</td>
<td>1696</td>
<td>1730</td>
<td>1772</td>
<td>1839</td>
</tr>
<tr>
<td>247</td>
<td>1634</td>
<td>1668</td>
<td>1722</td>
<td>1764</td>
<td>1819</td>
</tr>
<tr>
<td>278</td>
<td>1608</td>
<td>1640</td>
<td>1700</td>
<td>1746</td>
<td>1812</td>
</tr>
<tr>
<td>318</td>
<td>1574</td>
<td>1626</td>
<td>1669</td>
<td>1731</td>
<td>1785</td>
</tr>
<tr>
<td>371</td>
<td>1537</td>
<td>1591</td>
<td>1649</td>
<td>1695</td>
<td>1758</td>
</tr>
<tr>
<td>445</td>
<td>1504</td>
<td>1556</td>
<td>1605</td>
<td>1663</td>
<td>1718</td>
</tr>
<tr>
<td>556</td>
<td>1444</td>
<td>1506</td>
<td>1554</td>
<td>1603</td>
<td>1695</td>
</tr>
<tr>
<td>741</td>
<td>1382</td>
<td>1442</td>
<td>1485</td>
<td>1549</td>
<td>1671</td>
</tr>
<tr>
<td>1112</td>
<td>1290</td>
<td>1361</td>
<td>1403</td>
<td>1505</td>
<td>1632</td>
</tr>
<tr>
<td>% (171 veh/h over 741 veh/h)</td>
<td>23.5%</td>
<td>20.7%</td>
<td>19.5%</td>
<td>17.9%</td>
<td>11.8%</td>
</tr>
</tbody>
</table>

Table 7-2 presents the simulated merging capacity of the leftmost lane at different metering rate with different \( \theta \). Figure 7-4 demonstrates the changes of the merge capacity as the metering rate increasing. As can be seen clearly from Figure 7-4, there is a negatively relationship between the merge capacity and the metering rate. The lower the disturbance coefficient, \( \theta \), is, the higher impacts the disturbance has and the lower the merge capacity is. All the curves in Figure 7-4 can be divided into two parts, the part shown in the green rectangle and the one shown in the red rectangle. In the first part, the impact of each individual disturbance is basically independent of each other, so the curve is basically linear. In the second part, the
impact of each disturbance starts to overlap others, and the drop of the capacity is no longer linear with the increase of the metering rate.

![Merge Capacity](image)

**Figure 7-4 Merge capacity vs. metering rate**

Figure 7-5 is a comparison between the proposed model and an analytical model from Leclercq, Laval, et al. (2011). The y-axis in Figure 7-5 is the relative capacity drop, which is calculated as the reduced capacity divided by the theoretical capacity. As can be seen, they show the same trend: as the ramp flow increases, the relative capacity drop decreases, and $\theta=0.4$ gives the closest result from the proposed model to the analytical model.

![Comparison of estimated relative capacity drop](image)

(a) Results from the proposed model  
(b) Results from Leclercq, Laval, et al. (2011)

**Figure 7-5 Comparison of estimated relative capacity drop**
The simulation settings are similar with the Test-bed S of the micro-simulation analysis in Section 5.5.1. Compared with the micro-simulation results in Table 5-3, the percentage increase of the minimum metering rate (171 veh/h/l in Table 7-2) over the high metering rate 741 veh/h/l in Table 7-2 (1112 veh/h/l is over the maximum metering rate in Australia) is the closest when $\theta=0.4$ (23.5% in Table 7-2 and 22.4% in Table 5-3). Therefore, results of $\theta=0.4$ are selected for regression, and the relationship is used for evaluation in Section 7.4. The regression equation is as follows:

$$Q_{\text{merge}}(t) = 224.1 \cdot \ln(r(t)) + 2863.8$$  \hspace{1cm} (7-7)

where $r(t)$ is the metering rate.

### 7.4 EVALUATION

This section reports the evaluation results of the ZRM-RCR strategy using the M-CTM with the variable merging capacity.

#### 7.4.1 Simulation Network in CTM

An artificial network is built for the evaluation. In order to build a CTM model, the first step is to set up parameters for a cell. There are 5 basic parameters for each cell:

- Critical density $“k_c”$: the critical density where achieves the capacity, and 20 veh/km/lane is used in the simulation;
- Free flow speed $“Spd_f”$: the model sets 100 km/hour as the free flow speed;
- Capacity $“Q_c”$: a product of critical density and free flow speed, so $Q_c = k_c \times Spd_f = 2000$ veh/h/lane;
- Jam density $“k_j”$: the density where the motorway is totally filled by vehicles;
- Cell length $“l_{cell}”$: according to (Daganzo, 1994), the length of the cell is set equal to the distance travelled in light traffic by a typical vehicle in an update interval; in the model, the update interval is set as 30 seconds; therefore, $l_{cell} = Spd_f \times 30\text{seconds} = 833.3\text{meters}$. 


The next step is to design network geometry. In this simulation exercise, the network is designed to contain 4 on-ramps, and the most downstream on-ramp is the bottleneck where mainline queue originates (See Figure 7-6), and the mainline is a 3-lane motorway.

Figure 7-6 shows two mainline cells between on-ramps, 4 merge cells (C1, C4, C7 and C10), and 4 ramp cells. C0 is the most downstream cell, so it is set to have infinite capacity. The five source cells, C11 and the four ramp cells, are set to have infinite space for queuing vehicles; therefore, all the vehicles will enter the network once they are generated by the demand scenario.

As the simulation exercise is designed to simulate the congestion recovery scenario, the initial state is set as the colour shown in Figure 7-6. A mainline queue starts from C1 (the bottleneck) and propagates back to C5. Densities at C5 and C6 are close to critical density, and the mainline further upstream is totally free flow. Apart from the mainline, all the ramps hold medium on-ramp queues because it is assumed that the coordinated RM is running until this time point. R1 and R2 are designed as long ramp, with 130 veh as the maximum queue length and 80 veh as the initial queue length. The other two upstream on-ramps, R3 and R4, are short ramps, with 70 veh as the maximum queue length, while the initial queue length is 35 veh.

In terms of the demand profile, the mainline entrance demand and the on-ramp demands are decreasing in the congestion recovery scenario. Figure 7-7 illustrates the demand profile of the five sources for the 30-min simulation.

**7.4.2 Modelling RM in CTM**

In macroscopic models like CTM, the RM algorithm produces the number of vehicles from the on-ramp into the merge cell for each time interval. Based on the ramp incoming flow, the merge cell is then able to determine its interaction with the upstream mainline cell. Therefore, the RM algorithm is updated first at every update.
interval. Accordingly, the flow chart of CTM with RM algorithm is demonstrated in Figure 7-8.

*Figure 7-7 Demand profile*

Two different algorithms are modelled in this simulation exercise: the CRM, and the ZRM-RCR. Therefore, two test scenarios are simulated.

In the M-CTM, the merging capacity is related to ramp flow rate. For simplicity, only the capacity of the leftmost lane is assumed to be affected by the ramp traffic using Equation 7-7; the equation is given as follows:
\[ Q_{\text{merge}}(t) = 224.1 \cdot \ln(r(t)) + 6863.8 \quad (7-8) \]

### 7.4.3 Performance Indicators

The four performance indicators used in the evaluation as follows:

- **Total travel time (TTT)**: TTT is the most widely used efficiency indicator at a system level for RM, and it is calculated by summing up number of vehicles in all the cells through the whole simulation period and then multiple by update interval. Its unit is veh∙h;

- **Ramp total travel time (RTTT)**: RTTT is similar to TTT but only calculates the ramp cells. Its unit is also veh∙h;

- **Queue spillover time (QSOT)**: it is defined as the total time for one or all ramp cells when queue length is equal to or over the maximum queue length. The unit is minute;

- **Average mainline travel delay (AMTD)**: CTM is a macroscopic model, so it cannot calculate individual vehicle travel time and delay; also CTM does not provide speed information directly; therefore, cell speed is assumed to be a piecewise linear function of cell density (see Figure 7-9), and then cell speed is converted to cell travel time and delay. The sum of mainline cell delay is defined as mainline travel delay for each time interval. Then, AMTD is the time average of mainline travel delay. The equation of calculating cell speed is given as follows:

\[
Spd = \begin{cases} 
Spd_f (100 \text{ km/h}), & k \leq k_c \\
-1.5k + 130, & k_c < k < k_{\text{queue}} \\
Spd_{\text{queue}} (10 \text{ km/h}), & k \geq k_{\text{queue}}
\end{cases} \quad (7-9)
\]

### 7.4.4 Results

Table 7-3 shows the aggregated results while Table 7-4 gives RTTT and QSOT for each ramp. Figure 7-10 demonstrates the density contours of both test scenarios.

Overall, the ZRM-RCR strategy outperforms the CRM, as expected. One purpose of using CTM for evaluation is to crosscheck the concept of the ZRM-RCR. TTT is the performance indicator which is calculated to be almost the same as in Chapter 6. When compare the percentage reduction of TTT with its counterpart, TTT-R in Table 6-8, the magnitudes from both models are at the same level. In
addition, other performance indicators also show a similar trend. This confirms the conclusion made in Chapter 6.

![Figure 7-9 Piecewise linear relationship between speed and density](image)

Table 7-3 Simulation results summary using modified CTM

<table>
<thead>
<tr>
<th>Unit</th>
<th>CRM</th>
<th>ZRM-RCR</th>
<th>Reduction</th>
<th>Percentage reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>veh-h</td>
<td>416.7</td>
<td>396.1</td>
<td>20.6</td>
</tr>
<tr>
<td>RTTT</td>
<td>veh-h</td>
<td>119.2</td>
<td>98.7</td>
<td>20.5</td>
</tr>
<tr>
<td>QSOT</td>
<td>minute</td>
<td>9</td>
<td>8.5</td>
<td>0.5</td>
</tr>
<tr>
<td>AMTD</td>
<td>sec/trip</td>
<td>93.1</td>
<td>89.3</td>
<td>3.8</td>
</tr>
</tbody>
</table>

Table 7-4 Individual ramp results using modified CTM

<table>
<thead>
<tr>
<th>Ramp</th>
<th>RTTT</th>
<th>CRM</th>
<th>ZRM-RCR</th>
<th>QSOT</th>
<th>CRM</th>
<th>ZRM-RCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30.8</td>
<td>54.9</td>
<td>4</td>
<td>8.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>33.9</td>
<td>23.9</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>27.3</td>
<td>10.0</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>27.3</td>
<td>10.0</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>119.2</td>
<td>98.7</td>
<td>9</td>
<td>8.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results from Table 7-4, together with the density contours in Figure 7-10, clearly show how the ZRM-RCR obtains the benefit. By restricting on-ramps R1 and R2 in the mainline queuing area, the whole system recovers more quickly. The mainline queue can be seen to have been quickly reduced and restricted around the
merge bottleneck in the ZRM-RCR scenario. Although traffic from R1 experiences more delays, the time of the queue spillover at R1 does not increase seriously due to the reduced demand. This advises that the recovery phase is another chance for the system to gain benefit by RM with small system costs. Overall, the work mechanism and process shown with the M-CTM are similar to the results in Chapter 6. This further confirms the effectiveness of the proposed strategy.

**Figure 7-10 Mainline density contours – modified CTM**

### 7.5 SUMMARY

In this chapter, the proposed ZRM-RCR strategy is evaluated in an artificial network modelled by CTM. A hybrid modelling framework is proposed to construct the numerical relationship between the merging capacity and the metering rate. Then, the merge cell in the CTM is modified by applying variable merging capacity determined by the metering rate. The evaluation results confirm the effectiveness of the ZRM-RCR strategy. Moreover, the results from CTM provide an explanation of the mechanism of the ZRM-RCR similar to the one concluded from the micro-simulation evaluation.
Chapter 8 Discussion and Conclusion

This research project investigated the feasibility of achieving rapid recovery for motorway congestion using ramp metering (RM) as the management tool and developed a RM strategy for rapid congestion recovery (RCR), namely a zone-based RM strategy for RCR (ZRM-RCR). Its development began by comprehensively reviewing existing RM algorithms from a control system point of view. Then, in order to build a bench mark RM system of the current state of practice, two critical components were studied. The developed local queue management scheme and the basic local RM algorithm, ALINEA, were used to build the local RM sub-system. The developed coordination strategy uses a two-layer structure: at a higher layer, the coordination is formulated from a rule-based heuristic approach; at a lower layer, with the PID (Proportion-Integration-Differentiation) controller developed for the slave ramp, the coordination is able to provide quicker and more flexible reaction. After setting up the bench mark system, an investigation for the feasibility of RCR by RM was conducted. The investigation started by analysing the changes in traffic flow dynamics at the recovery phase, and a micro-simulation study proved that the restrictive metering control (RMC) is able to increase the merging capacity at the recovery phase. In the next step, this research developed the ZRM-RCR strategy, and evaluated its effectiveness using two different test-beds in two different recovery scenarios. Evaluation results concluded the effectiveness of the ZRM-RCR strategy, especially for merging bottlenecks. In order to further confirm the conclusion, a modified cell transmission model (M-CTM), a popular macroscopic model, was used to evaluate the proposed strategy, and the results reached a similar conclusion. These conclusions from two different simulation models give confidence to answer the research questions posed in Chapter 1.

8.1 RESPONSE TO RESEARCH QUESTIONS

What is the system benefit of RCR?

From the system point of view, the earlier the motorway network recovers to normal condition (free flow) from congestion, the higher the system efficiency is achieved (less total travel time). By analysing the traffic flow features at the recovery
phase (details in Section 5.3), the reduced traffic demands are seen to give an opportunity to benefit both mainline and ramp traffic with controllable ramp costs. For the mainline, the decreased demand means no more mainline queue accumulation and no risk of mainline queuing once recovered. For the ramp, the ramp demand is reducing at the recovery phase, which means ramp costs are not as severe as during peak hours. As the mainline traffic recovers, ramp traffic can easily enter the motorway without the risk of congestion. Ultimately, ramp traffic gains the benefit from RCR.

**Can RM assist or accelerate motorway system recovery?**

There are two reasons for using RM to achieve RCR. Firstly, restricting ramp traffic means giving priority to mainline traffic. In other words, more mainline queuing traffic can pass the bottleneck as less ramp traffic is entering. Secondly, theoretical analysis (see in Section 5.3) and micro-simulation investigation (see in Section 5.5) both indicate that RMC operation at the recovery phase can increase the merging bottleneck throughput. The “extra” merging bottleneck throughput will further accelerate the mainline queue discharge. As a result, RMC operation will accelerate motorway system recovery if there are any active merging bottlenecks.

**How can we identify the recovery phase after motorway mainline congestion?**

In Chapter 6, two different methods were developed to identify the recovery phase for the peak hour congestion scenario and the incident scenario. For the peak hour congestion scenario, both mainline and ramp incoming flow rates and their projections are monitored. Once the decreasing trend in these variables is confirmed by consecutive intervals, the recovery phase is identified. Details can be found in Section 6.2.4.

The incident scenario (serious accident for example) usually causes serious traffic jams at the mainline. Once the incident is cleared, that is the time to consider recovery. However, the mainline queue caused by a traffic jam stays almost still, and there is no benefit to limit ramp traffic when the merging area is totally congested. Therefore, the recovery phase for each ramp in the queuing area is when the traffic at its merging area starts to flow. Considering that the cleared incident will increase the capacity at the incident location significantly, the mainline queue will then start to be
flushed. As a result, a backward queue dissipation shockwave will be generated. Accordingly, the recovery phase for each ramp is when the queue dissipation shockwave propagates back to its merging area. See details in Section 6.5.2.

**How can we design strategy using RM to achieve RCR?**

The developed ZRM-RCR strategy is the answer to this question. In ZRM-RCR, a two-phase control (compulsory control phase and reactive control phase) algorithm is proposed. In the compulsory control phase, bottleneck ramps are forced to activate RMC for rapid mainline queue discharge; in the reactive control phase, ramp costs (including queue spillover and ramp queues) are considered and managed. Comprehensive evaluation in both micro-simulation and macro-simulation models indicated the effectiveness of the proposed strategy in both congestion and incident scenarios: that is, the strategy can accelerate the system recovery and manage the total ramp costs at the same time.

### 8.2 LIMITATIONS AND FUTURE RESEARCH

Three limitations of this research, if addressed, could further improve the research for motorway management.

1. The target of this research is to develop a field-applicable RM strategy. Therefore, field tests are of high interest to the author, especially for testing the impact of RMC on merging bottleneck throughput. With field tests, it will be possible to better adjust the parameters or even modify the logic of the ZRM-RCR.

2. As identified by the literature review in Chapter 2, one future trend for motorway management is to involve more ITS (Intelligent Transportation Systems) tools together to achieve a system optimum. For example, the latest information technologies, like GPS (Global Position System) and vehicular communication networks, are able to provide accurate information and in real time. Specifically, GPS provides real-time vehicle location information, while the vehicular communication system can deliver the information to other vehicles or authorities in real time. The information can be used for many purposes. Firstly, a density measure can be derived from individual vehicle locations. Secondly, the vehicular information can be used to monitor the trend of traffic flow so as to make
the recovery phase identification much easier. Another instance is to consider cooperation with other motorway management tools, such as variable speed limits (VSL).

3. To the author’s best knowledge, there is no effective solution for the motorway weaving bottleneck with field evaluation reported in the literature. For example, the most significant bottleneck northbound on the Pacific Motorway is the interchange of the Gateway Motorway, which is a weaving bottleneck caused by huge diverging traffic to the Gateway Motorway. The fundamental reason for weaving congestion is the disturbances caused by many compulsory lane-changes. With macroscopic management tools, like RM and VSL, the only solution is to reduce density at the weaving area significantly enough to provide enough space for lane changes. However, this requires reducing large amounts of traffic flow to the weaving area, and that is impossible during peak hours. As vehicle to vehicle and vehicle to infrastructure communication will be available in the near future, it would be an opportunity to develop microscopic control (for individual vehicles) to cooperate these compulsory lane-changes.

8.3 CONCLUDING REMARKS

The following conclusions can be drawn from this research:

1. Recovery phase is the phase when the capacity starts to match the total traffic demands again, either due to the reduction of demands in peak congestion recovery or due to the increase of capacity by incident clearance. This is another opportunity for RM to accommodate demand and capacity with controllable costs endured by ramp traffic (due to reduced ramp demand), especially for merging bottlenecks.

2. RMC, the basic RM operation for the recovery phase, can accelerate mainline queue discharge (extra increased capacity can be gained for a merging bottleneck).

3. Micro-simulation results show that the ZRM-RCR strategy can accelerate the system recovery and manage the total ramp costs at the same time.
4. For a metered ramp, the metering rate is the main factor that affects the merging capacity, especially during the recovery phase. The numerical relationship between the merging capacity and the metering rate can be obtained by a hybrid modelling framework based on moving bottleneck theory and linear car-following model. The M-CTM uses the numerical relationship to generate variable merging capacity at merge cells.

5. Evaluation results from the M-CTM confirm the effectiveness of ZRM-RCR strategy at a macroscopic level.
References


