STATE OF CALIFORNIA • DEPARTMENT OF TRANSPORTATION
TECHNICAL REPORT DOCUMENTATION PAGE
TR0003 (REV 10/98)

1. REPORT NUMBER	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
CA16-2808		

4. TITLE AND SUBTITLE	5. REPORT DATE
Performance Analysis and Control Design for On-ramp Metering of Merging Bottlenecks	09/01/2016
	6. PERFORMING ORGANIZATION CODE

7. AUTHOR	8. PERFORMING ORGANIZATION REPORT NO.
Wenlong Jin and Felipe Augusto de Souza	

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NUMBER
Institute of Transportation Studies University of California, Irvine 4038 Anteater Instruction & Research Building (AIRB) Irvine, CA 92697-3600	11. CONTRACT OR GRANT NUMBER IA 65A0529

12. SPONSORING AGENCY AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
California Department of Transportation Division of Research Innovation, and System Information MS-83 1727 30th Street Sacramento CA 95816	final report 4/20/2015 - 3/31/2016
	14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This research aims to analyze the performance and design the control parameters for on-ramp metering of congested merging bottlenecks. One important motivation of ramp metering, variable speed limit, and other centralized and decentralized traffic control strategies is to prevent capacity drop, which occurs at active bottlenecks.

The complex interplay among merging, lane-changing, and accelerating behaviors plays an important role in determining the performance of a congested merging area. Especially, once a merging bottleneck is activated, the discharging flow-rate can drop by 10% (about 800 vph on a four-lane freeway); such a capacity drop can lead to excessive traffic queues and stop-and-go traffic patterns and increase fuel consumption and GHG emissions. The objective of this research is to analyze the performance and design the control parameters for both pretimed and traffic-responsive on-ramp metering of congested merging bottlenecks. In this research we will (1) quantify the congestion mitigation effects of different ramp metering algorithms at an active merging bottleneck, (2) design control parameters for efficient and robust traffic responsive ramp metering algorithms, (3) identify demand patterns when ramp metering algorithms are effective, and (4) develop a set of simple decision-support tools for ramp metering with both kinematic wave models and microscopic simulations. The research will help Caltrans to make decisions on the necessity, priority, algorithm, and parameter tuning related to ramp metering.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Ramp metering; Merging; Lane-changing; Capacity drop; Congestion mitigation effects; Control design; Paramics simulation	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161

19. SECURITY CLASSIFICATION (of this report)	20. NUMBER OF PAGES	21. COST OF REPORT CHARGED
Unclassified	50	

Reproduction of completed page authorized.

DISCLAIMER STATEMENT

University of California Transportation Center

Performance Analysis and Control Design for On-ramp Metering of Merging Bottlenecks

Wenlong Jin
Ph.D., Associate Professor
Civil and Environmental Engineering
Institute of Transportation Studies
University of California, Irvine

Felipe Augusto de Souza
Graduate Student
Civil and Environmental Engineering
Institute of Transportation Studies
University of California, Irvine

2016

University of California
Transportation Center

ABSTRACT

This research aims to analyze the performance and design the control parameters for on-ramp metering of congested merging bottlenecks. One important motivation of ramp metering, variable speed limit, and other centralized and decentralized traffic control strategies is to prevent capacity drop, which occurs at active bottlenecks.

Locally, we analyze and design ramp metering of an isolated merging bottleneck. We use a simple link queue model to describe traffic dynamics with an ordinary differential equation combined with a capacity drop model. This enables us to establish important system properties such as equilibrium states, reachability, and closed loop stability when PI-ALINEA algorithm is applied. We identified the set of equilibrium states and we show that for specific demand patterns there could be two equilibrium states. We establish reachability conditions, defined as the capability of steer the system to an uncongested state, and show that it is dependent on the initial state so there could be a situation in which is possible to avoid congestion, but once it is congested it is not possible to dissipate. Using the PI-ALINEA algorithm, we studied the closed loop response and we show the stability range, that is, for any initial condition it can steer to and settle at the desired state. Using Cell Transmission Model simulations, we validate the reachability property and closed loop stability.

Globally, we analyze the performance of several bottlenecks. With an example for a freeway segment with two bottlenecks, we demonstrate that successful control of some active bottlenecks can worsen the overall road network's performance; this is a paradoxical behavior of traffic control systems similar to the Braess paradox, where additional links can worsen a system's performance. In particular, we show that the overall performance of the system in stationary states becomes worse after the successful control of the upstream active bottleneck, if the following two conditions are satisfied. First, before control, only the upstream bottleneck is active; i.e., a queue appears only upstream to this bottleneck. Second, after the successful control of the upstream bottleneck, the downstream bottleneck is activated by a larger discharging flow-rate from the upstream bottleneck, and the formed queue spills back to the former and congests the upstream bottleneck. We simplify the sufficient conditions to demonstrate that such a paradoxical behavior occurs under a wide range of conditions with respect to capacity drop magnitudes, turning ratios, and demand levels. Thus, without careful analysis of their system-wide effects, some local control strategies can be myopic; this motivates us to define the price of myopia to quantify the potential negative effects of myopic control strategies. With numerical simulations we verify that such a paradoxical behavior also occurs under dynamic and random conditions when the upstream bottleneck is controlled by a local feedback variable speed limit strategy. This study highlights the necessity of holistic analysis and control of a highly nonlinear traffic system.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

Chapter 1

Introduction

1.1 Background

On freeways, congestion usually occurs during peak hours [Schrank et~al. (2012)], causing accidents/incidents, delays, higher fuel consumption, and air pollution [Golob and Recker (2003)]. Ramp metering is one of the possible techniques to improve freeway performance, as demonstrated in field deployments (e.g, [Papageorgiou et~al. (1997)],[Levinson and Zhang (2006)]) and simulation studies (e.g., [Jin and Zhang (2001)], [Sun and Horowitz (2006)], [Papamichail et~al. (2010)]). There are two mechanisms that ramp metering helps to reduce total travel time.

1. The first is related to the queue spill back mechanism. When a queue starts from a bottleneck, propagates upstream and reaches an upstream off-ramp, vehicles leaving at that off-ramp are delayed due to congestion ahead [Papageorgiou and Kotsialos (2000)]. These vehicles exiting from off-ramp do not go through the bottleneck, but are nonetheless impacted. By holding vehicles on on-ramps, ramp metering might prevent, or at least postpone, the queue from reaching that off-ramp and consequently reducing the total travel time.
2. The second is related to the capacity drop phenomenon, a drop in the discharge flow rate of a merge bottleneck when its upstream section gets congested [Cassidy and Bertini (1999)] while its downstream section is uncongested. The magnitude of the drop varies, and a typical value is around 10% [Chung et~al. (2007)]. As the flow rate is lower, vehicles take longer to pass through the bottleneck, and the total travel time is increased. Again, by storing the vehicles on on-ramps, ramp metering might prevent or postpone the onset of the congestion and, therefore, capacity drop.

It is not an easy task to quantify the share of each mechanism on the overall improvement [Papageorgiou (1998)]. Few empirical studies have related the impact of ramp metering on capacity drop. An important study is reported in [Cassidy and Rudjanakanoknad (2005)] that shows that through a more restrictive metering rate, it is possible to recover the discharge flow rate to capacity on a isolated merge bottleneck. Simulation studies have considered the effect of the capacity drop implicitly or explicitly. Using second order models, in [Smaragdis et~al. (2004)] it was shown that variations of ALINEA were able to sustain a higher outflow for local control. Similar models have been used for coordinated control as in [Papamichail et~al. (2010), Kotsialos et~al. (2002)]. More recently, capacity drop was taken explicitly and integrated in model based controllers (see [Han et~al. (2015), Maggi et~al. (2015)]).

1.2 Local Ramp Metering

As for other control systems, it can be helpful to analytically study the effect of capacity drop on the system dynamics controlled by an on-ramp meter. Through these studies, system properties can be established in closed form solution helping us to understand important features of the system. This ultimately can be used for general guidelines such warranties of a deployment, parameters tuning of established algorithms, and provide insights to the design of new algorithms.

An interesting study on closed-loop ramp metering and its operating regime is reported in [Gomes and Horowitz (2003)]. Using the Cell Transmission Model [Daganzo (1995)], it was shown that different ramp metering algorithms can be analyzed from operation "modes" and its transitions. It was established controllability and observability with respect to detector placement. The analysis, the authors claim, suggests that ALINEA [Papageorgiou et al. (1991)] is a superior strategy compared to %-Occ.

The set of equilibrium states and their characteristics in a single freeway was studied in [Gomes et~al. (2008)]. It was shown that all equilibrium states leads to the same flow rate on the bottlenecks, but keeping those bottlenecks uncongested is beneficial as it diminishes the aforementioned queue spill back effect. They show that through ramp metering, it is possible to steer the system to an uncongested equilibrium state and therefore reducing delays.

Closed loop stability for ramp metering also has been the subject of recent research, in particular on ALINEA and its variations. Through linearization and Lyapunov theory, stability is established for PI-ALINEA in [Wang et~al. (2014)]. For PI-Controllers and a class of systems that local ramp metering fits in, stability was also derived [Karafyllis and Papageorgiou (2014)]. The aforementioned study [Gomes and Horowitz (2003)] also establishes stability range for ALINEA.

All of these studies have not considered the capacity drop. Its understanding is an essential step in order to analyze and design ramp metering algorithms especially for local ramp metering control. We attempt to fill some of the gaps by analytically studying essential open and closed loop (with PI-ALINEA control law) properties. This is enabled by using models that are simple and yet capable of reproducing essential traffic flow characteristics. A link queue model [Jin (2012)] is used for the traffic dynamics inside the merging segment. This model is an approximation of the LWR model and extends the cell demand and supply functions to a link. Second, a simple model is incorporated to replicate the outcome of the capacity drop phenomenon at a merge bottleneck [Jin et~al. (2015)], that is, a decrease in flow on onset of congestion. The combined model lead to a switched linear ordinary differential equation [Liberzon (2012)]. We were able to establish the following:

1. The system has hysterical nature with respect to the demand pattern. Demand higher than the capacity triggers the congestion; however, in order to clear a formed congestion, it is necessary a demand lower than the current, and lower, capacity;

2. Depending on the amount of the capacity drop, it might not be possible for a local ramp controller to be effective in order to eliminate the congestion. In this situation, while the meter can at some extent change the proportion of the delays on on-ramp and mainline freeway, it is not able to clear the congestion; and

3. If the controller is able to effectively eliminate the congestion, the stability region of the widely applied and studied (PI-)ALINEA algorithm is derived.

1.3 Global Ramp Metering

As we know, transportation systems are highly nonlinear, and naive strategies may not always lead to better system performance. A sublime example is the occurrence of the Braes paradox [Braess et~al. (2005)], which reveals that the addition of a link can deteriorates the whole system's performance. Then one may wonder whether successful control of one or more active bottlenecks is always beneficial to the whole system. In this study, we present one example to show that it may not be the case. That is, we demonstrate the existence of a paradoxical behavior in traffic control systems, where the successful control of individual active bottlenecks could actually deteriorate the overall system's performance. Such a paradoxical behavior of traffic control systems can lead to wrong investment decisions and worse traffic congestion and therefore warrants a better understanding.

Here we consider a simple example in a freeway corridor with two bottlenecks. We demonstrate that, in stationary states, the overall performance of the system can be worsened after the introduction of a local control for the upstream bottleneck, if the network satisfies the following conditions before and after control. Before control, only the upstream bottleneck is active, and the downstream bottleneck is uncongested; that is, congestion only develops at the upstream bottleneck before control. Then a naive solution is to develop a local control strategy for the upstream bottleneck. For example, we can apply the variable speed limit (VSL) strategy developed in [Jin and Jin (2015)]. However, we can show that the system's total discharging flow-rate is smaller, if after control the downstream bottleneck is activated by a larger discharging flow-rate from the upstream bottleneck, and the formed queue spills back to the former. With the help of a macroscopic capacity drop model [Jin et~al. (2015)], we demonstrate that such a paradoxical behavior occurs under a wide range of conditions with respect to capacity drop magnitudes, turning ratios, and demand levels. That is, under the sufficient conditions such a VSL strategy is myopic, and we further compare the total travel times before and after the myopic control and define the price of myopia to quantify its negative effects. With dynamical and random demand patterns, we carry out numerical simulations to examine the existence of such a paradoxical behavior and calculate the corresponding price of myopia.

Chapter 2

Reachability and Stability for Local Ramp Metering System

The system under study is depicted in Figure 2.1 and referred to as a merge bottleneck, containing three components: the merge gore between the on-ramp and the freeway at $x = 0$, the bottleneck (lane drop) located downstream at $x = L$, where the on-ramp acceleration lane ends, and the merging segment between the merge and the bottleneck.
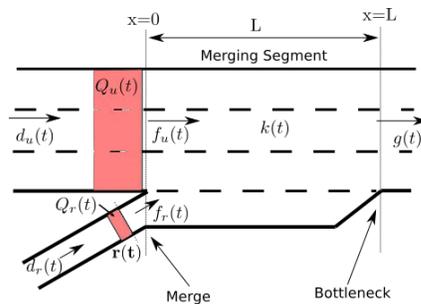
Figure 2.1: Schematic of merge bottleneck and the model variables.

The state variable is the average density in the merging segment, $k(t)$. The inputs include the mainline demand, $d_u(t)$, and the on-ramp demand, $d_r(t)$. The total demand is denoted by $d(t) = d_u(t) + d_r(t)$. In addition, $f_u(t)$ is the mainline in-flux, $f_r(t)$ the on-ramp in-flux, and $g(t)$ the out-flux. The metering rate is denoted by $r(t)$. On the upstream boundary and on-ramp any unserved traffic is modeled as point queues [Jin (2015)], where $Q_u(t)$ and $Q_r(t)$ are the respective queue sizes.

2.1 System Description and Model

The traffic dynamics inside the merging segment can be described by the Lighthill-Whitham-Richards (LWR) model [Lighthill and Whitham (1955), Richards (1956)], which has been successfully applied to analyze the initialization, propagation, and dissipation of traffic congestion with spatial and temporal density waves (kinematic waves). However, the LWR model is a partial differential equation, more specifically a hyperbolic conservation law, for which the control problem is not well studied. In this study, we resort to an approximate model, the Link-Queue Model (LQM) [Jin (2012)], which only considers dynamical variations of spatially average densities and is therefore an ordinary differential equation. In [Jin and Jin (2014)], this model has been successfully applied to analyze and design the variable speed limit strategy, and the results are validated in the LWR model thorough Cell Transmission Model

simulation. Thus we follow the same approach by studying the control of merge bottlenecks with the LQM:

$$\dot{k}(t) = \frac{1}{L}(f(t) - g(t)), \tag{1}$$

where $f(t)$ and $g(t)$ are the in and out-fluxes. Equation (1) can be viewed as a reservoir in which level increases or decreases based on the in- and out-fluxes difference. The fluxes are computed based on demand and supply concepts [Daganzo (1995), Lebacque (1996)]

$$D(t) = \min\{v_f k(t), v_f k_c\} \tag{2}$$

$$S(t) = \min\{(v_f k_c, \omega(k_j - k(t))\}, \tag{3}$$

which are respectively the increasing and decreasing parts of the triangular fundamental diagram [Newell (1993)]:

$$q(k) = \min\{v_f k, \omega(k_j - k)\}, \tag{4}$$

where $q(k)$ is the flow-rate, v_f the free flow speed, k_j the jam density, and ω the shock-wave speed. The density which yields maximum flow is $k_c = \dfrac{\omega k_j}{v_f + \omega}$; at this point the flow is the capacity, $C = v_f k_c$. Ramp, acceleration lane, and freeway lanes share the same characteristics, as v_f and ω. We denote k_{jl}, k_{cl} and C_l as per-lane jam-density, critical density, and capacity. Note that $k_j = nk_{jl}$, $k_c = nk_{cl}$ and $C_n = nC_l$, where n is the number of lanes. Also, hereafter C refers to downstream capacity unless stated otherwise. The density that yields capacity is $k = k_{cd}$. Figure 2.2 presents the fundamental diagram at each segment.
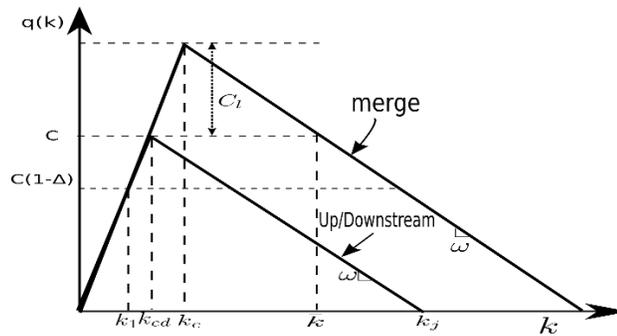
Figure 2.2: Fundamental diagram of upstream, downstream, and merging segments.

On the upstream and on-ramp unserved vehicle are modeled as point queues:

$$\dot{Q}_u(t) = d_u(t) - f_u(t) \tag{5}$$

$$\dot{Q}_r(t) = d_r(t) - f_r(t), \tag{6}$$

where the demands are computed [Jin (2015)]:

$$D_u(t) = \min\left\{\frac{Q_u(t)}{\epsilon} + d_u(t), v_f k_c\right\}, \tag{7}$$

$$D_r(t) = \min\left\{\frac{Q_r(t)}{\epsilon} + d_r(t), C_r\right\}, \tag{8}$$

where $\epsilon = \lim_{\Delta t \to 0^+} \Delta t$.

The on-ramp vehicles are assumed to have absolute priority, and the on-ramp flux is given by:

$$f_r(t) = \min\{D_r(t), S(t), r(t)\}, \tag{9}$$

the remaining supply can be used for the upstream flow

$$f_u(t) = \min\{D_u(t), S(t) - f_r(t)\}, \tag{10}$$

and the total in-flux is $f(t) = f_u(t) + f_r(t)$. We also denote as $D_m(t) = D_u(t) + D_r(t)$ as the total demand on the merge. Note that $f(t) = \min\{D_m(t), S(t)\}$.

At the downstream boundary of the merging segment, the out-flux is determined by:

$$g(t) = \min\{D(t), C(1 - \Delta H(k(t) - k_{cd}))\}, \tag{11}$$

where $H(x)$ is the Heaviside function:

$$H(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{12}$$

and Δ is the capacity drop ratio.

It is assumed that there is no congestion on the downstream mainline freeway, i.e., the merge bottleneck is active. Note that the capacity drop model proposed in [Jin et~al. (2015)] is used here to replicate the capacity drop phenomenon: when there is no queue on the merging segment, the out-flux can reach the downstream capacity, but if a queue forms, the out-flux is the

dropped capacity $g(t) = C(1-\Delta)$. An important aspect is that the drop ratio, Δ, is exogenous and should be determined for each case.

2.2 Why Ramp Metering? The Impact of the Capacity Drop on Delay

From the presented model, it is possible to assess the impact of capacity drop by comparing the drop and not drop case. It is assumed that any transitory period is small compared to the total time considered. We consider two cases, one discharging at capacity, C, and another at dropped capacity, $C(1-\Delta)$. The total demand, $d = d_u + d_r = \alpha C$ is assumed to be over capacity (i.e., $\alpha > 1$) by $t = T$ and zero thereafter.

In Figure 2.3 cumulative curves $N(t)$ are depicted. The continuous line is the cumulative arrival. The dashed lines are the departure rates for the case in which it discharges at capacity, C and the case discharging at dropped capacity, $C(1-\Delta)$. The vertical difference between the arrival and departure curve is the instantaneous queue. This queue could be at on-ramp, mainline, or both depending on the upstream and ramp demand and the metering rates.
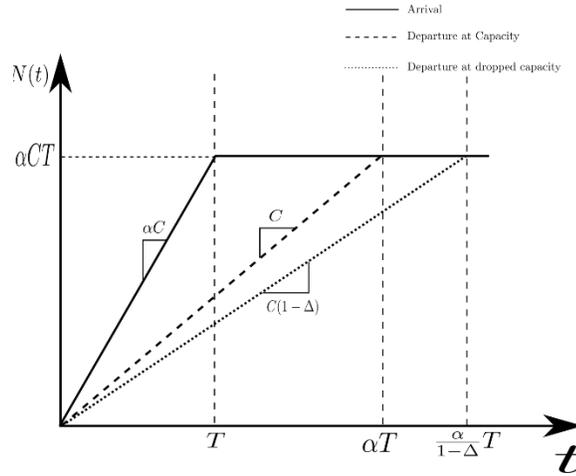
Figure 2.3: Arrival (continuous) and departure (dashed) cumulative curves. The outflow (departure) could be either at capacity, C, or at dropped capacity, $C(1-\Delta)$.

The areas between the arrival and departure curve are the total delay:

$$D_{nd} = \frac{1}{2}CT^2[(\alpha-1)\alpha]$$

(13)

$$D_d = \frac{1}{2}CT^2[\frac{(\alpha+\Delta-1)\alpha}{1-\Delta}]$$

(14)

The relative improvement of avoiding the capacity drop is given by:

$$D(\%) = 1 - \frac{D_{nd}}{D_d} = \frac{\alpha_1(1-\overline{\Delta})}{\alpha_1 - \overline{\Delta}} = \frac{\alpha_1 \Delta}{\alpha + \Delta - 1} \qquad (15)$$

For example, if $\alpha = 1.1$ (that is, demand 10 % higher than capacity) and $\Delta = 0.05$, the improvement is 36%; if the drop amount $\Delta = 0.1$, the difference goes to 55%. Therefore, a well designed ramp meter can drastically decrease the delay. The question turns to which conditions should be satisfied to achieve such reduction.

2.3 When Ramp Metering Is Effective? The Equilibrium States and Reachability Property

The ramp meter is effective if it can lead the system to a desired state. In particular, as the delay is lower when the out-flux is higher, the goal is to discharge at capacity. First, we show the equilibrium states and its characteristics. It is shown that keeping at uncongested equilibrium states is beneficial. Then, we show in which conditions it is possible to lead the system to the uncongested equilibrium state.

2.3.1 The equilibrium states and their Behavior

We analyze the equilibrium states of the systems subject to constant metering rate ($r = C_r$ for no control case) and ignoring the on-ramp and mainline queues. In this case the total demand is constant:

$$D_m = \hat{d} = d_u + \min(r, dr). \qquad (16)$$

The system reaches equilibrium states classified as follows:

- If $\hat{d} < C(1-\Delta)$, the system reaches an uncongested equilibrium density $k_{eq} = \dfrac{\hat{d}}{v_f} < k_1$

from any initial state.

- If $C(1-\Delta) \le \hat{d} \le C$, the system reaches an uncongested equilibrium density $k_{eq} = \dfrac{\hat{d}}{v_f} \in [k_1, k_{cd}]$ from an initial state $k(0) \le k_{cd}$.

- If $C(1-\Delta) \le \hat{d} \le C$, the system reaches a congested equilibrium density $k_{eq} = \overline{k}$ from an initial state $k(0) > k_{cd}$.

- If $\hat{d} > C$, the system reaches a congested equilibrium density $k_{eq} = \overline{k}$ from any initial state.
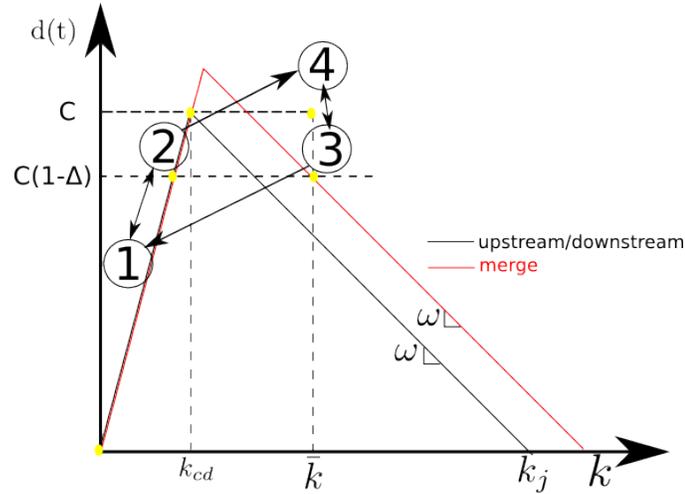
Figure 2.4: Transition in equilibrium states subject to the change in the demand level. Yellow dots highlight the boundaries between states. Uncongested states are 1 and 2; congested states are 3 and 4.

These states and transition between states are represented in Figure 2.4. The system shifts to state 1 whenever $\hat{d} < C(1-\Delta)$ either initially at state 2 or 3. Similarly, reaches state 4 when $\hat{d} > C$ either from states 2 or 3. However, it shifts to state 2 when $C(1-\Delta) \leq \hat{d} \leq C$ and initially at state 1. Likewise, reaches state 3 for the same demand level, but initially on state 4.

Even though the model is based on continuous variables, its essential operating regimes and its transitions can be characterized by a finite state machine. Note that a complete cycle can be done clockwise, but it is not possible on the other way around. This fact shows the inherent hysteresis: when the system is initially at state 2, it is necessary a demand greater than capacity ($\hat{d} > C$) to reach state 3; however, it is necessary demands lower than the dropped capacity (i.e., $\hat{d} < C(1-\Delta)$) to return to state 1 again and then a demand to $C(1-\Delta) \leq \hat{d} \leq C$. It is not possible to switch between state 3 and 2 without state 1 as intermediate.

It is clear that keeping in state 2 has advantages over state 3. In state 2 yields higher out-flux while keep the bottleneck uncongested. However, to shift state 3 to 2 is not straightforward. As it needs to shift to state 1 first, the system needs a sharp reduction in demand in order to recover the capacity.

The hysteretical nature of transportation networks have been discussed with empirical evidence in [Geroliminis and Sun (2011)]. However, the capacity drop phenomenon it happens on a single merge and is not outcome of the capacity drop phenomenon and not queue spill back.

2.3.2 Equilibrium States Classification

The equilibrium state can be characterized over different aspects regarding its equilibrium. We analyze for convergence and stability.

Under constant demand, the system is convergent [Gomes et~al.(2008)Gomes, Horowitz, Kurzhanskiy, Varaiya, and Kwon]: given constant demand d, it always converges to one equilibrium state, either the $k_{eq,u} = \dfrac{\hat{d}}{v_f}$ or congested $k_{eq,c} = \bar{k}$. Any density in the interval $(k_{cd}, \bar{k})$ is an unstable equilibrium state for $\hat{d} = C(1 - \Delta)$.

For stability, we analyze based on Lyapunov stability [Astrom and Murray(2010)] in which an equilibrium state is stable if the initial condition is close to an equilibrium, it will remain close to this equilibrium. In this case, the small perturbation could be either at the state variable, density, or in the demand.

With respect to demand level there are two cases in which it fails. It initially at critical density and $\hat{d} = C$, a demand $\hat{d} + \gamma$ where γ is small and greater than zero will lead the system to the congested equilibrium $\bar{k}$. Likewise, if $k(0) > k_{cd}$ and $\hat{d} = C(1 - \Delta)$; a demand $\hat{d} - \gamma$ will lead the system to $k = \dfrac{\hat{d} - \gamma}{v_f} << \bar{k}$. In this sense, it can be classified as bistable: the system has two distinct equilibrium points depending on the sign of the perturbation on the demand.

In this case a perturbed demand lead to distinct equilibrium states, but the same is true for a small perturbation on density. If $C(1 - \Delta) \leq \hat{d} \leq C$ and $k = k_{cd}$. A small perturbation positive on density leads to $\bar{k}$ while a negative leads to $k_{eq,u} = \dfrac{d}{v_f}$.

Figure 2.5 depicts the bifurcation diagram considering d as bifurcation parameter [Wiggins (2003)]. Continuous line represents stable equilibrium and dashed lines unstable equilibrium. Bifurcations have been discussed for traffic network in [Daganzo et~al. (2011)] and [Jin (2013)]. The existence of multiple stationary states in a network in [Jin (2015)] and multiple equilibrium states in a single freeway as in [Gomes et~al. (2008)] implies bifurcations. However, all of them the underlying principle is a queue spill back effect which means that flows reduces on upstream links or sections due to a congestion downstream. In this case, the possibility of multiple equilibrium states arise in a single merge and affects also the downstream flow.

As mentioned and it also can be seen on the bifurcation diagram that $k(0) > k_{cd}$ and $C(1 - \Delta)$ is an unstable equilibrium state. The fact it is unstable it does not change the operation

regimes in 4 as the stable equilibrium states are the ones likelier to be observed in practice [Daganzo et~al. (2011)] and therefore more important to be studied.
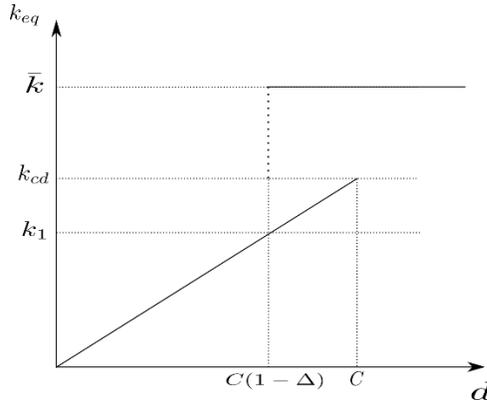
Figure 2.5: Bifurcation diagram: the set of equilibrium states for varying demands.

2.3.3 Reachability with dynamic metering rates

For demands between the capacity and the dropped capacity, the system can be either congested or uncongested. In order to shift it is needed a demand lower than the dropped capacity for a sustained period. As the metering rate can limit the on-ramp demand, we ask the following question: in which condition the ramp meter is able to avoid the congestion? If initially congested, in which condition is it possible to dissipate the congestion?

We use the terminology of control theory in which reachability is the capability to reach an arbitrary state [1] through any function $r(t)$ [Astrom and Murray (2010)] that satisfies the constraints (this case $r_{min} \leq r(t) \leq C_l$). When there exists at least one $r(t)$ that satisfies this condition, the state it is reachable. For unconstrained linear systems, a general test is often used [Kalman et~al. (1960)]. However, the system under analysis is switched and $r(t)$ constrained.

In this case, the goal is to keep the system uncongested and therefore discharging at capacity. Then, the test consists in keeping $k(t) \leq k_{cd}$. Let Z the set of points such $k(t) \leq k_{cd}$, then Z is reachable if the controller is able to lead the system to at least one element of Z. We assume here constant demands, either upstream, $d_u(t) = d_u$, or ramp, $d_r = (t) = d_r$.

Theorem 2.3.1 Z *is reachable for* $d_u < C(1-\Delta) - r_{min}$ *for any initial condition.*

---

[1] The term controllability and reachability are often exchangeable depending on the textbook [Astrom and Murray (2010)], here we follow the definition as in [Sun et~al.(2002)Sun, Ge, and Lee]: controllability is related to reaching the origin and reachability is related to an arbitrary state.

Proof. If $k(t_0) > k_1$, $g(t) = C(1-\Delta)$. Setting $r(t) = r_{min}$, $f(t) = \min\{v_f k_c, \omega(k_j - k(t), d_u + r_{min}\}$. As long as $d_u < C(1-\Delta) - r_{min}$, $f(t) < C(1-\Delta)$ and $\dot{k}(t) < 0$ and eventually, at $t = t_1$, $k \le k_{cd}$. Once it is uncongested, either initially or at $t = t_1$, it will remain if $r(t) \le C - d_u$ because $\dot{k}(t) = \frac{1}{L}(f(t) - \min(C, v_f k(t))$ and $g(t) \ge f(t)$ at boundary, $k = k_{cd}$, and $\dot{k}(t) \le 0$ so $k(t) \le k_{cd}$ for $t > t_1$. $+$

Theorem 2.3.2 Z *is reachable for* $d_u < C - r_{min}$ *if* $k(t_0) < k_{cd}$

Proof. With $r(t) = C - r_{min}$, follows the same condition of Theorem 4.1 for $t > t_1$.

$+$

Outside this region, the controller is no longer effective. For example, if the system is in State 3 and $d_u > C - (1-\Delta) - r_{min}$, even with $r(t) = r_{min}$ it does not switch to any of the uncongested states (1 or 2). In this case, a drop in the upstream demand, d_u, is necessary to relieve the congestion.

Though still able to control $k(t)$, there might exist unserved demand. In order to keep at full capacity, the metering rate can be set to levels lower than ramp demand forming a queue that may spillover to local streets.

Theorem 2.3.3 *For* Z *reachable, all demand is served for* $d < C$.

Proof. If Z is reachable implies that eventually $k(t) \le k_{cd}$ and it can operate at capacity. After this instant, $r(t) = C - d_u(t)$ can be set. The maximum influx is $f(t) = d_u + \min(r(t), C - d_u$ and $f(t) \le d_u(t) + C - d_u$, thus $f(t) \le C$. If $d = d_u + d_r < C$, then $d_r < C - d_u$, so $r(t) \ge d_r$ and both upstream and ramp demand are served. $+$

When demand exceeds capacity, queues will grow either on on-ramp or upstream. Case $d_u + d_r > C$, setting the metering rate as $r = C - d_u$ the upstream and ramp flux would be $f_u = d_u$ and $f_r = r$ respectively. The on-ramp queue would evolve as:

$$\dot{Q}_r = d_r(t) - f_r(t) = d_r - r = d_u + d_r - C \ge 0, \tag{17}$$

The queue would steadily increase. In practice, this queue has a maximum length in order to avoid the congestion to spill over to local streets. Often, the meter has a queue override feature that forces a higher metering rate to avoid long queues on on-ramps. It is not considered

explicitly here. However, at this point either delay will increase on local streets, due to queue spill back, or at mainline due to the capacity drop (see (15) .

When reachability is not guaranteed, the control system is able, at some extent, to change the share of the delays on on-ramp or mainline freeway, but it will discharge at dropped capacity.

This result, while in this case for a single merge, differs from [Gomes et~al.(2008)Gomes, Horowitz, Kurzhanskiy, Varaiya, and Kwon]. It was shown that there could be multiple equilibrium states for a bottleneck with demand larger than capacity. As the capacity drop phenomenon was not considered, the resulting flow rate at the bottleneck is unique and always at capacity. Considering the capacity drop phenomenon, the flow rate is lower when the bottleneck is congested.

Also on that study, it was proven that it is possible to *steer* the system towards the uncongested equilibrium state. It is a similar to what we defined on this study for reachability. While in the modeling considered here, it was introduced a minimum metering rate r_{min}, even for $r = r_{min} = 0$ if $d_u > C(1-\Delta)$ it is not possible to dissipate the congestion.

On the case of coordinated control, d_u is function of metering rates on upstream on-ramps on previous time. Clearly, if $r_{min} = 0$ on all on-ramps it is possible to induce a $d_u = 0 < C(1-\Delta)$. However, if the flow induced by $r_m in > 0$ on all upstream on-ramps lies in the interval $(C(1-\Delta), C)$ it is possible to avoid the capacity drop, but not recover from it.

This also shows the impact of the minimum metering rate. A higher minimum metering rates can make Z not reachable. For this purpose, ideally $r_{min} = 0$; however, usually agencies might impose higher minimum metering rates due to other operational issues, such as Caltrans in California [Sun and Horowitz(2005)].

2.4 How to Lead the System to the Desired State? Closed-Loop Analysis and Parameter Design

The model equations are combined with PI-ALINEA [Wang et~al. (2010)] in order to analyze the response in closed-loop. First, the PI-ALINEA algorithm is briefly described. Second, the choice of set-point, k_o, and the equilibrium states are discussed. Then, we show for which parameters, K_p and K_i, the system in closed loop is stable. A Poincaré map analysis is presented for the case which the response is oscillatory.

2.4.1 PI-ALINEA

ALINEA [Papageorgiou (1991)] is a feedback control algorithm based on PID Controller family. The metering rate is updated based on the observed occupancy just before the lane-drop. While the traditional ALINEA is an I-controller, in this study we consider the extended PI-ALINEA [Wang et~al. (2010)], which considers a PI-Controller rather than an I-Controller. Also, the ALINEA control law is considered in discrete time. In this study we consider the continuous PI-Controller, given by [Astrom and Murray (2010)]:

$$r(t) = K_p e(t) + K_i \int_0^t e(\tau)d\tau, \tag{18}$$

where K_p and K_i are the proportional and integral coefficients respectively, and the error, $e(t)$, is the difference between the real-time density $k(t)$ and the target density $k_o(t)$:

$$e(t) = k_o(t) - k(t). \tag{19}$$

In addition, the control signal $r(t)$ is bounded:

$$r_{min} \leq r(t) \leq C_l. \tag{20}$$

Thus, it is necessary to determine the following parameters: the coefficients K_p and K_i, the target density $k_o(t)$, and the minimum metering rate r_{min}.

2.4.2 Set-Point Specification and Equilibrium States

From the analysis of equilibrium states, the fundamental diagram, and the PI-ALINEA control law (Eq. 18), we can find the optimal set point:

1. if $d > C$ the maximum out-flux is when $k(t) = k_{cd}$ with out-flux $g(t) = C$;

2. if $d \leq C$ and $k(t) \geq k_{cd}$ the system could have been operating with the same out-flux, but at free-flow speed ($k \leq k_{cd}$). In this case any set point less or equal to k_{cd} will force the freeway to operate at free-flow speed; and

3. if $d \leq C$ and $k(t) \leq k_{cd}$ the system will not operate at capacity; however, any control action in the direction of a higher flux (i.e., increase of metering rate) is always desirable or at least does not affect the performance. The integral effect will push the metering rate to $r(t) = C_l$ as long as the set point, k_o, is such that $k(t) < k_o \leq k_{cd}$.

Therefore, the set point $k_o = k_{cd}$ always leads the system to its maximum throughput in steady state, considering no fluctuation in demand or modeling errors.

The equilibrium states depend on the PI-controller set-point and demand. From Equation (18) , the PI-Controller holds constant metering rate $r(t)$ when $e(t) = 0$, as long as $K_i > 0$. In other words, the PI-Controller assures that the only equilibrium point is $k(t) = k_o = k_{cd}$. However, this state may not be reachable depending on demand patterns.

If the demand is high enough to not match the reachability condition (see Section 2.3.2), it is not possible to avoid the congestion and capacity drop. Indeed, when $k(t) > k_{cd}$, $r(t)$ will steadily decrease until the lower bound $r(t) = r_{min}$. Once $d_u(t) > C(1-\Delta) - r_{min}$, the influx will be higher than out-flux until $k(t) = \overline{k}$, which is the equilibrium state in this case. Once Z is reachable again (i.e., $d_u < C(1-\Delta) - r_{min}$), the density will start do decrease and eventually $k(t) \le k_{cd}$.

On the other hand, the set-point might not be reached for low demands. For $k(t) < k_{cd}$ the metering rate will steadily increase until the upper bound $r(t) = C_l$, but as long as $d(t) < C$ it is always possible to serve the demand and the equilibrium state is $k(t) = \dfrac{d(t)}{v_f}$.

2.4.3 Closed Loop Response and Stability

For stability analysis, we assume:

- Z is reachable so that there is a $r(t)$ that can lead to the set point;

- the ramp flux is determined by the metering rate (i.e., $f_r(t) = \min(r(t), d_r(t)) = r(t)$). The on-ramp queue evolve as $\dot{Q}_r(t) = d_r(t) - r(t)$ and will not be included in the state space;

- the upstream queue is ignored and assumed to be zero and $f_u(t) = d_u(t)$. In this case it is implicitly assumed that $d_u(t)$ reflect any queue or unserved vehicles up to time t.

Also, the upstream demand is split in a constant and a variable term: $d_u(t) = d_{u0} + \delta(t)$ and we define excess density as $x(t) = k(t) - k_{cd}$ and excess demand as $v(t) = r(t) + d_{u0} - C$. The on-ramp queue is not included on the state variable as it is a direct consequence of the assumption .

Combining equations (1) , (9) , (11) , and the control law (18) , the system can be described as:

$$\dot{x}(t) = \frac{1}{L}(r(t) + d_{u0} + \delta(t) - v_f x(t) - C) \tag{21}$$

$$\dot{r}(t) = -\frac{K_p}{L}(r(t) + d_{u0} + \delta(t) - v_f x(t) - C) - K_i x(t), \tag{22}$$

which is valid for $x(t) \le 0$. Similarly, for $x(t) > 0$:

$$\dot{x}(t) = \frac{1}{L}(r(t) + d_{u0} + \delta(t) - C(1 - \Delta)) \tag{23}$$

$$\dot{r}(t) = -\frac{K_p}{L}(r(t) + d_{u0} + \delta(t) - C(1 - \Delta)) - K_i x(t). \tag{24}$$

Setting excess density and demand as the state variables, $Y(t) = [x(t), v(t)]^T$, we have the following switched affine system [Lin and Antsaklis (2009)]:

$$\dot{Y} = A_1 Y + B_1 + P\delta(t), x(t) \le 0, \tag{25}$$

$$\dot{Y} = A_2 Y + B_2 + P\delta(t), x(t) > 0, \tag{26}$$

where

$$A_1 = \begin{bmatrix} -\dfrac{v_f}{L} & \dfrac{1}{L} \\ \dfrac{K_p}{L} v_f - K_i & -\dfrac{K_p}{L} \end{bmatrix}, \ B_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ P = \begin{bmatrix} \dfrac{1}{L} \\ -\dfrac{K_p}{L} \end{bmatrix},$$

$$A_2 = \begin{bmatrix} 0 & \dfrac{1}{L} \\ -K_i & -\dfrac{K_p}{L} \end{bmatrix}, \ \text{and} \ B_2 = \begin{bmatrix} \dfrac{C\Delta}{L} \\ -K_p \dfrac{C\Delta}{L} \end{bmatrix}.$$

The demand variation is treated as perturbation, $\delta(t)$, which is assumed zero throughout this analysis. The equilibrium points are $Y_1^{\mathring{a}}(t) = [0,0]^T$ for (25) and $Y_2^{\mathring{a}}(t) = [0, -C\Delta]^T$ for (26). In both cases $x = 0$ is the equilibrium point; however, $Y_1^{\mathring{a}}$ is ideal because the out-flux is greater.

Locally, the dynamic is determined by the eigenvalues of the respective matrix A_k ($k = 1,2$):

$$\lambda(A_1) = \sigma_{1n} \pm j\omega_1 = \frac{-\frac{v_f + K_p}{L} \pm \sqrt{\frac{(v_f + K_p)^2}{L^2} - \frac{-4K_i}{L}}}{2}, \tag{27}$$

$$\lambda(A_2) = \sigma_{2n} \pm j\omega_2 = \frac{-\frac{K_p}{L} \pm \sqrt{\frac{K_p^2}{L^2} - \frac{-4K_i}{L}}}{2}, \tag{28}$$

where $\lambda(A_k)$ denotes the eigenvalues of matrix A_k.

Through the nature of the eigenvalues in respect to the sign of its real part and whether it is a complex number and the initial condition after switching between regimes, the stability is derived. The complete derivation is in the Appendix I.

The stability is guaranteed by two basic condition. After switching from congested to uncongested state, it remains uncongested and converges to $Y_1^å$. Also, if initially congested it should be guaranteed that the system will eventually switch to the uncongested state.

The first condition is guaranteed by real and negative eigenvalues. While it is possible a switch to the congested state depending on the initial conditions, an eventual transition back to the uncongested state always will be $Y(0) = [0, -C\Delta + \varepsilon]$, $\varepsilon > 0$ and for this initial condition real and negative eigenvalues will lead the system to the origin.

The second condition, is to assure that the system initially congested eventually switches to the uncongested regime. In this case, a real and negative eigenvalue can settle the system on the congested side. Complex eigenvalues guarantees a transition. In the specific case where eigenvalues are real and positive, the system always switches back as long as $K_i > 0$ due to the saturation. As the real part is positive, the system initially diverges from $x = 0$ and reaches $r(t) = C_l$; at this point $K_p x(t)$ keeps constant while the integral term increases; when the integral term exceeds the proportional, the system is pushed to the congested side.

Combining all these cases, the system is stable and converge to $Y_1^å$ when:

$$K_p > -v_f \tag{29}$$

$$H(K_p)\frac{K_p^2}{4L} \leq K_i \leq \frac{(v_f + K_p)^2}{4L} \tag{30}$$

An interesting fact is that the drop amount, Δ, does not influence the eigenvalues. So, these results would be the same as long as the drop amount is greater than zero.

Another addressed case is when eigenvalues of A_1 and A_2 are all complex numbers. In this case, whatever the initial conditions, it has $x(t) = a_n sin(\omega_n t)$ and it always cross the line $x = 0$ at $\dfrac{\pi}{\omega_n}$. In the new region, it changes ω_n and a_n, but not the functional form and it always switches back after half a period. We use Poincaré Map [Wiggins(2003)] analysis to analyze the behavior of the oscillations over multiple cycles.

Consider the Figure 2.6, we assume the initial condition at $Y(0) = [0, v_1]^T$. It follows a sinusoidal trajectory and intercepts again $x = 0$ in the point $Y_2 = [0, v_2]$. This process repeats until point $Y_3 = [0, v_3]^T$ and so on. At each segment, $v_i = f_1(v_{i-1})$ and therefore $v_i = f_2(v_{i-2})$. After obtaining $f_2(v)$ it is possible to compute when it will cross the segment $x = 0$ coming from the same dynamic region after n cycles and what is the asymptotic behavior when $n \to \infty$.

With the response given by Equation 2 and (48) and (49), v_2 and v_3 are obtained:

$$v_2 = Y(\frac{\pi}{\omega_1}) = \frac{v_1}{\omega_1 L} e^{\frac{\pi\sigma_1}{\omega_1}},$$ (31)

$$v_3 = Y(\frac{\pi}{\omega_1} + \frac{\pi}{\omega_2}) = (\frac{v_2 + C\Delta}{\omega_1 L}) e^{\frac{\pi\sigma_2}{\omega_2}}.$$ (32)

Combining both equations:

$$v_3 = v_1 e^{\frac{\pi\sigma_1}{\omega_1} + \frac{\pi\sigma_2}{\omega_2}} + C\Delta e^{\frac{\pi\sigma_2}{\omega_2}}.$$ (33)

Equation (33) has one fixed point ($v_1 = v_3 = v^{\mathring{a}}$) at $v^{\mathring{a}} = \dfrac{C\Delta e^{\frac{\pi\sigma_2}{\omega_2}}}{1 - e^{\frac{\pi\sigma_1}{\omega_1} + \frac{\pi\sigma_2}{\omega_2}}}$, configuring a stable limit cycle. Also note that if $\dfrac{\pi\sigma_1}{\omega_1} + \dfrac{\pi\sigma_2}{\omega_2}$ is positive, the trajectories will increase over time; when negative it asymptotically goes to $v^{\mathring{a}}$. The period is $\dfrac{\pi}{\omega_1} + \dfrac{\pi}{\omega_2}$.
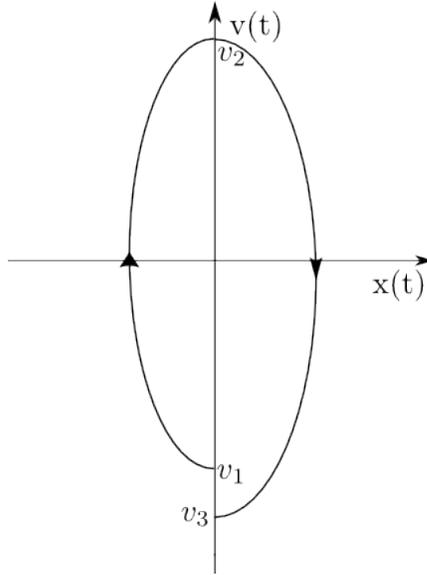
Figure 2.6: Phase diagram $x(t) \times v(t)$ with initial condition in $Y(0) = [0, v_1]^T$.

On the other hand, for $\dfrac{\pi\sigma_1}{\omega_1} + \dfrac{\pi\sigma_2}{\omega_2}$ is negative, the system will approach $v = v^{\mathring{a}}$ as $t \to \infty$.

Chapter 3

Paradoxical Behavior of Global Traffic Control Systems
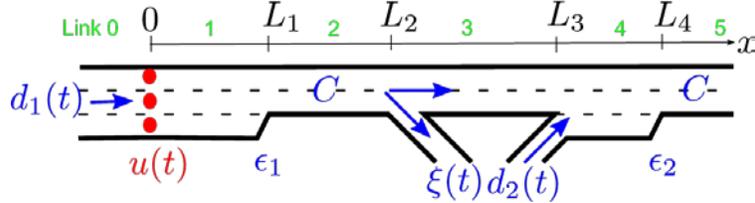
Figure 3.1: A freeway network

In this study, we consider a freeway network illustrated in Figure 3.1 , where there is an upstream lane-drop bottleneck at L_1, a diverge between the mainline freeway and an off-ramp at L_2, a merge between the mainline freeway and an on-ramp at L_3, and a downstream lane-drop bottleneck at L_4. The two lane-drop bottlenecks, the diverge, and the merge divide the freeway into six links, which are labeled from upstream to downstream as links $0,1,\cdots,5$. The capacities of links 2, 3, and 5 are assumed to be the same as C, which is smaller than the capacities of links 0, 1, and 4. Here we assume that link 5 is always uncongested; i.e., no queues will spill back from the further downstream part of the network.

The freeway demand is denoted by $d_1(t)$, and the on-ramp demand by $d_2(t)$. At the diverge, the turning ratio of vehicles to the off-ramp is denoted by $\xi(t) \in [0,1]$.

3.1 Capacity Drop and Variable Speed Limit Control

3.1.1 Capacity drop and stationary states at lane-drop bottlenecks

It is well known that, if traffic is congested upstream to a lane-drop bottleneck at either L_1 or L_4, the discharging flow-rate is smaller than the downstream capacity [Banks (1991), Hall and Agyemang-Duah (1991)]. Generally, the maximum discharging flow-rate for such a bottleneck can reach up to 2300 vphpl in free-flow traffic [Federal highway administration (1985), Hall and Agyemang-Duah (1991)], but when the upstream link is congested, the discharging flow-rate can be dropped by a magnitude in the order of 10% [Persaud et~al. (1998) Cassidy and Bertini (1999), Bertini and Leal (2005), Chung et~al. (2007)]. The capacity drop magnitude of an isolated active bottleneck is relatively stable for a location, but fluctuations in

discharging flow-rates can be caused by interactions among several bottlenecks [Kim and Cassidy (2012)].

In the literature various mechanisms have been proposed to explain the occurrence of capacity drop. For examples, an acceleration zone can be observed around the bottleneck [Banks (1991)], and capacity drop can be a consequence of the way drivers accelerate away from the upstream queue [Hall and Agyemang-Duah (1991), Papageorgiou et~al. (2008)]; it could be caused by an extensive queue on the shoulder lane upstream to the merging point, sharp declines in vehicle speeds, and increases in lane-changing activities [Cassidy and Rudjanakanoknad (2005)].

It is well known that, in the Cell Transmission Model [Daganzo (1995)], the upstream traffic becomes congested when the upstream demand exceeds the downstream supply. In [Jin et~al. (2015)], a simple kinematic wave model with a discontinuous boundary flux function in terms of the upstream demand and the downstream supply was proposed to explain capacity drop at the macroscopic level. In this model, capacity drop immediately occurs when the upstream demand is greater than the downstream supply, and the capacity drop magnitudes are given exogenously. As shown in Figure 3.1 , the capacity drop magnitudes at the two lane-drop bottlenecks are denoted by ε_1 and ε_2 respectively. Here $1 \geq \varepsilon_1 \geq 0$ and $1 \geq \varepsilon_2 \geq 0$: a capacity drop magnitude of zero means no capacity drop.

In [Jin et~al. (2015)], the macroscopic capacity drop model was shown to yield stationary states consistent with observations. In particular, a lane-drop bottleneck with capacity drop can be stationary in three traffic states, as shown in Figure 3.2 (a). First, the bottleneck is uncongested, if both upstream and downstream links are uncongested (stationary states are 1-1' in the figure); in this case, the discharging flow-rate is determined by the upstream link's demand and can be as high as the downstream link's capacity. Second, it is active, if the upstream link is congested, but the downstream link uncongested (stationary states are 2-2' in the figure); in this case, the discharging flow-rate always equals the dropped capacity. Third, it is congested, if both upstream and downstream links are congested (stationary states are 3-3' in the figure); in this case, a queue spills back from the downstream link, and the discharging flow-rate is determined by the downstream link's supply and cannot be higher than the dropped capacity.
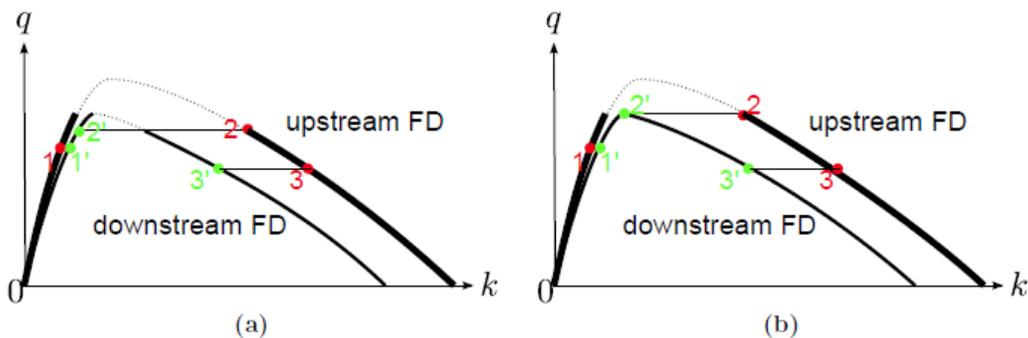

(a)   (b)

Figure 3.2: Stationary states at a lane-drop bottleneck: (a) Before control; (b) After control

3.2.2 Variable speed limit control

For the freeway network shown in Figure 3.1, under certain situations it is possible that traffic is congested upstream to L_1 but uncongested downstream. That is, the upstream bottleneck at L_1 is active [Daganzo (1999)], but the downstream one at L_4 uncongested. In this case, capacity drop occurs at L_1, and the discharging flow-rate is only $(1 - \varepsilon_1)C$, which is smaller than the capacity of link 2.

Naively, we would introduce a control strategy for the active upstream lane-drop bottleneck, so that the discharging flow-rate can be increased to C. In addition, since the downstream bottleneck is not active initially, one would think that there is no need to introduce any control for this location.

For example, variable speed limits (VSL) can be effectively implemented to improve the throughput at the upstream lane-drop bottleneck [Greenberg and Daou (1960), Cascetta et~al. (2011), Carlson et~al. (2011), Carlson et~al. (2013)]. Theoretically, for the network in Figure 3.1, where $u(t)$ denotes the variable speed limit upstream to the upstream lane-drop bottleneck, capacity drop at the upstream bottleneck can be totally prevented, since a VSL strategy can completely control the upstream demand. [2]

Therefore, after successful control of the upstream bottleneck, the three types of stationary states are shown in Figure 3.2 (b). Compared with those before control, the active state becomes critical, in which the discharging flow-rate equals the downstream link's capacity. Note that, after control, the upstream states are for those on link 0, not link 1, where transitional states exist due to the VSL control.

However, such local control strategies usually do not consider the impacts on other parts of the network. Thus they can be myopic if they worsen a system's performance. We refer to this phenomenon as a paradoxical behavior of a traffic system, as at the first glance one would expect that successful control of an active bottleneck should improve the system's performance. Another important reason for the occurrence of such a paradoxical behavior is that such strategies are helpless when the downstream part of a bottleneck is congested; i.e., by just managing the upstream demand, they cannot effectively alleviate traffic congestion when the downstream part is congested.

---

[2] In reality the upstream lane-drop bottleneck can also be a part of a merge bottleneck as the downstream bottleneck. In this case, integrated control strategies based on both VSL and ramp metering can be implemented to prevent the occurrence of capacity drop [Hegyi et~al.(2005)Hegyi, De~Schutter, and Hellendoorn, Zhang et~al.(2006)Zhang, Chang, and Ioannou, Allaby et~al.(2006)Allaby, Hellinga, and Bullock, Lu et~al.(2010)Lu, Varaiya, Horowitz, Su, and Shladover, Carlson et~al.(2010)Carlson, Papamichail, Papageorgiou, and Messmer]. But in this study for the purpose of simplicity we just consider a lane-drop bottleneck at L_1.

3.2 Sufficient Conditions for the Existence of the Paradoxical Behavior in Stationary States and Price of Myopia

In this section, we assume that the demands, $d_1(t) = d_1$ and $d_2(t) = d_2$, and the turning ratio, $\xi(t) = \xi$ are all time-independent. We present a set of sufficient conditions for the occurrence of a paradoxical behavior in stationary states, when traffic conditions are time-independent on all links. By comparing the system's performance before and after successful upstream control, we demonstrate that such a local control is myopic under these conditions. Without loss of generality, we assume the road is initially empty.

3.2.1 Sufficient conditions before and after control

In this subsection, we determine conditions with respect to capacity drop magnitudes, ε_1 and ε_2, turning ratios, ξ, and demand levels, d_1 and d_2, such that the network satisfies the following conditions:

• Before control, only the upstream bottleneck is active, and the downstream bottleneck is uncongested; that is, congestion only develops at the upstream bottleneck.

• After control, the upstream bottleneck's capacity drop is prevented, but the downstream bottleneck is activated, and the formed queue further congests the upstream bottleneck.

Before control, if the upstream bottleneck is active in stationary states, then the freeway demand d_1 has to be high enough:

$$d_1 > C, \tag{1}$$

and the discharging flow-rate at L_1 equals the dropped capacity, $(1-\varepsilon_1)C$. Therefore, the off-ramp flow-rate equals $(1-\varepsilon_1)\xi C$, and link 3's stationary flow-rate equals $(1-\varepsilon_1)(1-\xi)C$. If the downstream bottleneck is uncongested, then the total demand at the merge cannot be higher than link 5's capacity; i.e.,

$$(1-\varepsilon_1)(1-\xi)C + d_2 \le C. \tag{2}$$

It is straightforward to verify that, when both (1) and (2) are satisfied, the upstream bottleneck is active, and the downstream one uncongested in stationary states.

After successful control of the upstream bottleneck, then the capacity drop is effectively prevented for the bottleneck, and its discharging flow-rate is increased to C. If $(1-\xi)C+d_2 > C$; i.e., if [3]

$$d_2 > \xi C, \tag{3}$$

then the demand for the downstream bottleneck exceeds link 5's capacity, and the bottleneck is activated with a discharging flow-rate of $(1-\varepsilon_2)C$. Once the downstream bottleneck becomes active, a queue forms on link 4 and spills back to link 3. Assuming that on-ramp vehicles have the absolute priority to merge at L_3 [Daganzo(1995), Jin(2010)] and the on-ramp demand, d_2, is not larger than the dropped capacity of the downstream bottleneck; i.e.,

$$d_2 < (1-\varepsilon_2)C, \tag{4}$$

then link 3's stationary flow-rate is $(1-\varepsilon_2)C-d_2 > 0$. The queue further spills back to link 2, whose stationary flow-rate equals $\dfrac{(1-\varepsilon_2)C-d_2}{1-\xi}$, assuming that vehicles do not change their routes and the turning ratio is still ξ. In this case, the discharging flow-rate of the upstream bottleneck is determined by link 2's flow-rate, which is smaller than C from (3). If we assume that the discharging flow-rate is smaller than the dropped capacity; i.e., $\dfrac{(1-\varepsilon_2)C-d_2}{1-\xi} < (1-\varepsilon_1)C$, which leads to

$$(1-\varepsilon_1)(1-\xi)C+d_2 > (1-\varepsilon_2)C, \tag{5}$$

then the upstream bottleneck becomes congested in stationary states. It is also straightforward to verify that, given (3), (4), and (5), the network satisfies the aforementioned condition after control.

Lemma 3.2.1 *Conditions (2) - (5) are equivalent to the following three sets of conditions: (i) there exists capacity drop at both bottlenecks:*

$$1 > \varepsilon_1 > 0, \tag{6}$$

$$1 > \varepsilon_2 > 0; \tag{7}$$

(ii) the turning ratio is not too large:

---

[3] Otherwise, if (3) is not satisfied, then after successful control of the upstream bottleneck both bottlenecks become uncongested in stationary states.

$$0 \le \xi < 1 - \varepsilon_2; \tag{8}$$

(iii) the on-ramp demand is in a reasonable range:

$$d_2 > \max\{\xi C, (1 - \varepsilon_2)C - (1 - \varepsilon_1)(1 - \xi)C\}, \tag{9}$$

$$d_2 < (1 - \varepsilon_2)C, \tag{10}$$

$$d_2 \le C - (1 - \varepsilon_1)(1 - \xi)C. \tag{11}$$

Proof. From conditions (2) - (5) we find that the on-ramp demand, d_2, has two upper bounds: $C - (1 - \varepsilon_1)(1 - \xi)C$ and $(1 - \varepsilon_2)C$, and two lower bounds: ξC and $(1 - \varepsilon_2)C - (1 - \varepsilon_1)(1 - \xi)C$. For d_2 to exist, the two lower bounds have to be smaller than the two upper bounds. Therefore we can derive (3.1) and (8). Once these conditions are satisified, the range of q_2 is given by (3.1).   +

Note that the requirements on both capacity drop magnitudes in (3.1) and the turning ratio in (8) are quite mild. For example, there is no need for the downstream bottleneck's capacity drop magnitude to be larger, and relatively small turning ratios to the off-ramp can also be easily achieved in reality. Among the conditions on the demand levels, (1) is easy to be satisfied, but the on-ramp demand d_2 needs to be in a reasonable range, defined in (3.1). For example, on a two-lane free-way, $C \approx 4000$ vph, if we let $\varepsilon_1 = \varepsilon_2 = \xi = 0.1$, then the range of the on-ramp demand is $400 < d_2 \le 760$ vph, which is reasonably large. Therefore, (1), (8), (3.1), and (3.1) represent a quite wide range of conditions.

3.2.2 Existence of paradoxical behavior and price of myopia

When the network satisfies the two conditions before and after control in the preceding subsection, both of the two bottlenecks end up with worse stationary states after successful control of the upstream bottleneck, since the state of the upstream bottleneck changes from active to congested, and that of the downstream bottleneck from uncongested to active. Therefore, the total discharging flow-rate of the network is expected to be smaller after control, and the corresponding travel time in the system is larger. This leads to the paradoxical behavior that successful control of an active bottleneck worsens the system performance. In this subsection, we demonstrate the existence of such a paradoxical behavior and quantify the negative effects of such myopic local control strategies, under the sufficient conditions given by (1), (8), (3.1), and (3.1).

Before control, the upstream bottleneck is active, and its discharging flow-rate is $(1 - \varepsilon_1)C$, the off-ramp flow-rate is $(1 - \varepsilon_1)\xi C$, and link 3's stationary flow-rate is

$(1-\varepsilon_1)(1-\xi)C$; the downstream bottleneck is uncongested, and its discharging flow-rate equals $d_2 +(1-\varepsilon_1)(1-\xi)C$. Thus the network's total discharging flow-rate before control equals the off-ramp flow-rate plus the downstream bottleneck's discharging flow-rate, which is the upstream bottleneck's discharging flow-rate plus the on-ramp demand

$$g = (1-\varepsilon_1)C + d_2. \tag{12}$$

In contrast, after control, the network's total discharging flow-rate equals the upstream bottleneck's discharging flow-rate plus the on-ramp demand

$$\hat{g} = \frac{(1-\varepsilon_2)C - d_2}{1-\xi} + d_2 = \frac{1-\varepsilon_2}{1-\xi}C - \frac{\xi}{1-\xi}d_2. \tag{13}$$

Theorem 3.2.2 *Under (1), (3.1), (8), and (3.1), successful control of the upstream bottleneck leads to smaller total discharging flow-rate of the network; i.e.,*

$$\hat{g} < g. \tag{14}$$

That is, successful local control actually leads to worse system performance.

Proof. The conclusion is guaranteed by (5) or (3.1a). Note that, even though other conditions in (1), (3.1), (8), and (3.1) are not directly used in the proof of (14), they are essential for calculating both g and $\hat{g}$. One can verify that the off-ramp flow-rate and the discharging flow-rate of the downstream bottleneck are both smaller after control. In this sense, (1), (3.1), (8), and (3.1) are sufficient conditions for the existence of the paradoxical behavior in the network.   +

The deterioration of the system's performance is solely caused by the VSL control law or other local control methods, which simply aim to maximize the throughput of the upstream active bottleneck. Therefore such strategies are myopic and impose a price of myopia. During a period of T, we denote the system's total travel times before and after control by τ and $\hat{\tau}$, respectively. Then we define the price of myopia as the ratio of the increased travel time to the original travel time; i.e.,

$$\theta = \frac{\hat{\tau}}{\tau} - 1, \tag{15}$$

which is similar to the price of anarchy caused by selfish route choice behavior that leads to the Braess paradox [Roughgarden (2005)]. Therefore, when $\theta > 0$, a local control strategy has negative effects on the overall system.

In stationary states, when both freeway and on-ramp demands, d_1 and d_2, are constant, and traffic conditions in the network in Figure 3.1 are stationary during T. Then the total queueing times (excluding the free-flow travel times) before and after control are respectively

$$\tau = \frac{1}{2}T^2 \cdot \left(\frac{d_1+d_2}{g}-1\right),$$

$$\hat{\tau} = \frac{1}{2}T^2 \cdot \left(\frac{d_1+d_2}{\hat{g}}-1\right).$$

Thus the price of myopia in stationary states is

$$\theta = \frac{d_1+d_2}{d_1-(1-\varepsilon_1)C} \cdot \frac{(1-\varepsilon_1)(1-\xi)C+d_2-(1-\varepsilon_2)C}{(1-\varepsilon_2)C-\xi d_2}. \tag{16}$$

For example, $d_1 = 1.5C$, $\varepsilon_1 = \varepsilon_2 = \xi = 0.1$, and $d_2 = 0.15C$, then $g = 1.05C$, and $\hat{g} = 0.9833C$; and the price of myopia is $\theta = 0.19$. In this case, the myopic control increases the total queueing time by 19%.

Theorem 3.2.3 *For the network in Figure 3.1 with constant demands, successful control of the upstream bottleneck leads to a positive price of myopia in (16), when (1), (3.1), (8), and (3.1) are satisfied. In addition, θ increases in both d_2 and ε_2 but decreases in d_1 and ε_1. Therefore, its upper- and lower-bounds independent of the demands can be calculated in terms of ε_1, ε_2, and ξ as in the following:*

$$\theta < \frac{(1-\varepsilon_1)(2-\varepsilon_2)}{\varepsilon_1(1-\varepsilon_2)}, \quad \theta \le \frac{\varepsilon_2}{\varepsilon_1}\frac{2-(1-\varepsilon_1)(1-\xi)}{1-\varepsilon_2-\xi+(1-\varepsilon_1)(1-\xi)\xi}, \tag{17}$$

$$\theta > \max\{\frac{\varepsilon_2-\varepsilon_1+\varepsilon_1\xi}{1-\varepsilon_2-\xi^2},0\}. \tag{18}$$

where ε_1, ε_2, and ξ satisfy (3.1) and (8).

Proof. From (16), when (1), (3.1), (8), and (3.1), we can show that $d_1-(1-\varepsilon_1)C>0$, $(1-\varepsilon_2)C-\xi d_2 >0$, and $(1-\varepsilon_1)(1-\xi)C+d_2-(1-\varepsilon_2)C>0$. Thus $\theta>0$ in (16). It is also straightforward to show that θ increases in both d_2 and ε_2 but decreases in d_1 and ε_1. Furthermore, since θ increases in d_2 and decreases in d_1, it reaches the upper-bound when $d_1 = C$ and $d_2 = \min\{(1-\varepsilon_2)C,C-(1-\varepsilon_1)(1-\xi)C\}$, and the lower-bound when $d_1 = \infty$ and $d_2 = \max\{\xi C,(1-\varepsilon_2)C-(1-\varepsilon_1)(1-\xi)C\}$. Hence we obtain the range of θ given in (3.3).

3.3 Paradoxical Behavior under Dynamic and Random Conditions

Constant demands and parameters allow for straightforward proof of the existence of the paradoxical behavior and calculation of the price of myopia in stationary states, as shown in the preceding section. In this section, we demonstrate that such a paradoxical behavior also exists with dynamic demand levels and random capacity drop magnitudes and turning ratios, as in the real world.

For the same network in Figure 3.1 , we assume that all links have the same length of 600 m. The numbers of lanes for links 1 to 5 are 3, 2, 2, 3, and 2, respectively. We use the following triangular fundamental diagram [Munjal et~al. (1971), Haberman (1977), Newell (1993)]:

$$Q(a(x), \rho) = \min\{v_f \rho, w(a(x)k_j - \rho)\}, \tag{19}$$

where ρ is the traffic density, and $a(x)$ is the number of lanes at x. Here the parameters are from [Yang et~al. (2011)]: the free-flow speed $v_f = 30$ m/s, the shock wave speed in congested traffic $w = 35/8$ m/s, the jam density $k_j = 1/7$ veh/m/lane. Then the critical density is $k_c = \dfrac{w}{v_f + w} k_j = 1/55$ veh/m/lane, and the capacity of links 2 and 3 is $C = 2k_c v_f = 12/11$ veh/s for two lanes. The fundamental diagrams for $a = 2$ and $a = 3$ are shown in Figure 3.3.
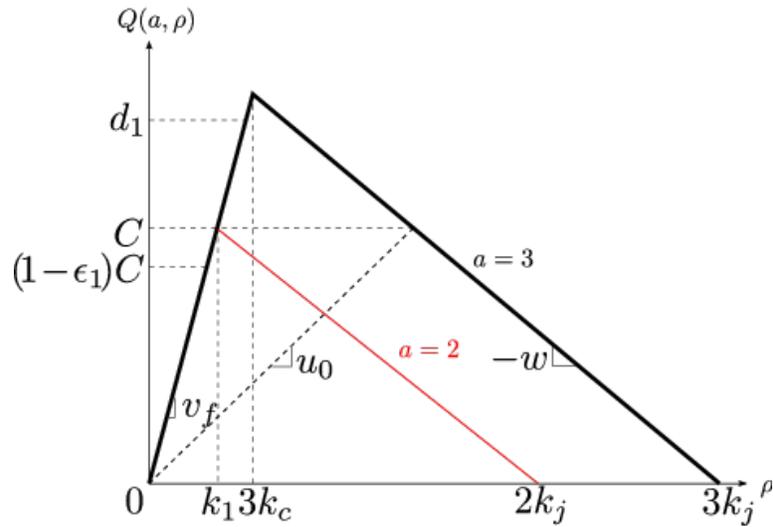

Figure 3.3: Fundamental diagrams

We assume that the capacity drop magnitudes and the turning ratio have an average value of 0.1 and independently follow a uniform distribution between 0.095 and 0.105. Their average values of these parameters satisfy (3.1) and (8). We choose $d_2(t)$, such that its average value during the peak period also satisfies (3.1):

$$0.1C < \bar{d}_2 \leq 0.19C. \tag{20}$$

We set $d_1(t) = \max\{0, \min\{\frac{t}{900}, 1, \frac{4500-t}{900}\}\} \cdot \bar{d}_1$, which simulates a peak-period demand pattern during an hour. In addition, we let the average mainline freeway demand during the peak period, $\bar{d}_1 = 1.2C$, and multiply $d_1(t)$ by a uniform random variable between 0.95 and 1.05 to add randomness to the demand pattern. Similarly, we set $d_2(t) = \max\{0, \min\{\frac{t-900}{900}, 1, \frac{4500-t}{900}\}\} \cdot \bar{d}_2$, which starts later than $d_1(t)$. We also multiply $d_2(t)$ by a uniform random variable between 0.95 and 1.05 to add randomness to the demand pattern. In our simulations, the network is empty initially.

3.3.1 Simulation set-up

In this study, we just simulate traffic dynamics on links 1, 2, 3, and 4, without considering link 5, which is always uncongested. We first discretize each link into $n = 20$ cells with the cell length of $\Delta x = \frac{600}{n} = 30$ m and a corresponding time-step size $\Delta t = 1$ s. We apply the Cell Transmission Model [Daganzo (1995)] to simulate traffic dynamics on all cells ($i = 1, \cdots, 4n$):

$$\rho_i^{j+1} = \rho_i^j + \frac{\Delta t}{\Delta x}(\mu_i^j - v_i^j), \tag{21}$$

$$\delta_i^j = Q(a_i, \min\{\rho_i^j, a_i k_c\}), \tag{22}$$

$$\sigma_i^j = Q(a_i, \max\{\rho_i^j, a_i k_c\}), \tag{23}$$

$$\mu_i^j = v_{i-1}^j = \min\{\delta_{i-1}^j, \sigma_i^j\}, \quad i \neq 1, n+1, 2n+1, 3n+1 \tag{24}$$

where the CFL number $v_f \frac{\Delta t}{\Delta x} \leq 1$ [Courant et~al.(1928)Courant, Friedrichs, and Lewy], ρ_i^j is the average density in cell i at $j\Delta t$, a_i the number of lanes in cell i, μ_i^j the in-flux of cell i

during $[j\Delta t, (j+1)\Delta t]$, v_i^j the out-flux of cell i during $[j\Delta t, (j+1)\Delta t]$, δ_i^j the demand of cell i at $j\Delta t$, and σ_i^j the supply of cell i at $j\Delta t$.

In addition, we assume that there are two point queues with infinite holding capacities at the mainline freeway entrance (at $x = 0$) and the on-ramp (at $x = L_3$). The queue lengths at $j\Delta t$ are denoted by λ_1^j and λ_2^j, respectively. Initially $\lambda_1^0 = \lambda_2^0 = 0$. We denote their demands by $\widetilde{d}_1^j$ and $\widetilde{d}_2^j$, respectively. Then from a standard point queue model [Jin (2015)], we have

$$\widetilde{d}_1^j = d_1^j + \frac{\lambda_1^j}{\Delta t}, \tag{25}$$

$$\widetilde{d}_2^j = d_2^j + \frac{\lambda_2^j}{\Delta t}, \tag{26}$$

where d_1^j is the mainline freeway demand at $j\Delta t$, and d_2^j is the on-ramp demand at $j\Delta t$.

At the upstream boundary of cell 1, we apply a PI-controller to determine the variable speed limit and calculate the in-flux as follows [Jin and Jin (2015)]:

$$u^j = u^{j-1} - \alpha(\rho_n^j - \rho_n^{j-1}) + \beta(k_1 - \rho_n^j)\Delta t, \tag{27}$$

$$u^j = \max\{u_{min}, \min\{u^j, v_f\}\}, \tag{28}$$

$$\mu_1^j = \min\{\widetilde{d}_1^j, \sigma_1^j, \frac{u^j w}{u^j + w} a_1 k_j\}, \tag{29}$$

where u^j is the variable speed limit at $j\Delta t$, $u_{min} = 0.5$ m/s is the minimum variable speed limit, $\alpha = 0$ and $\beta = 4$ are the coefficients of the PI-controller, and $k_1 = 2k_c$ is the target density which leads to the maximum discharging flow-rate at the upstream bottleneck. Here the VSL strategy adds an additional constraint on the boundary flux: $\frac{u^j w}{u^j + w} a_1 k_j$; when $u^j = v_f$, the constraint equals the capacity of cell 1, and it is equivalent to the case without upstream control. Thus it is reasonable to set the initial VSL as $u^0 = v_f$. Note that here the variable speed limit is adjusted based on the feedback of traffic conditions in cell n, which is the last cell in link 1. With such a local feedback scheme and the target density, this VSL strategy is myopic, without considering the potential impacts on the downstream bottleneck. k_1 and the corresponding speed limit, u_0, are illustrated in Figure 3.3.

At the lane-drop bottleneck between cells n and $n+1$, we apply the capacity drop model in [Jin et~al. (2015)] to calculate the boundary fluxes:

$$v_n^j = \mu_{n+1}^j = \min\{\delta_n^j, \sigma_{n+1}^j, (1-\varepsilon_1 \cdot H(\delta_n^j - \sigma_{n+1}^j))C\}, \tag{30}$$

where $H(y)$ is the Heaviside function:

$$H(y) = \begin{cases} 0, & y \le 0 \\ 1, & y > 0 \end{cases}$$

That is, when $\delta_n^j > \sigma_{n+1}^j$, the maximum discharging flow-rate is dropped to $(1-\varepsilon_1)C$; otherwise, the maximum discharging flow-rate is C. But the boundary flux is also bounded by the downstream supply, when link 2 is congested. This is the dynamic version of the capacity drop model described in Section 3.1, and the queue spillback effect from link 2 to link 1 is also captured.

At the diverge between cells $2n$ and $2n+1$, we apply the first-in-first-out diverge model in [Daganzo (1995)] to calculate the boundary fluxes:

$$v_{2n}^j = \min\{\delta_{2n}^j, \frac{\sigma_{2n+1}^j}{1-\xi^j}\}, \tag{31}$$

$$\mu_{2n+1}^j = v_{2n}^j(1-\xi^j), \tag{32}$$

$$q_{off}^j = v_{2n}^j \xi^j, \tag{33}$$

where ξ^j is the turning ratio at $j\Delta t$, and q_{off}^j the out-flux at the off-ramp. Here we assume that the off-ramp is always uncongested.

At the merge between cells $3n$ and $3n+1$, we apply the priority-based merge model in [Jin (2010)] to calculate the boundary fluxes:

$$\mu_{3n+1}^j = \min\{\delta_{3n}^j + d_2^j, \sigma_{3n+1}^j\}, \tag{34}$$

$$v_{3n}^j = \min\{\delta_{3n}^j, \max\{0, \sigma_{3n+1}^j - \tilde{d}_2^j\}\}, \tag{35}$$

$$q_{on}^j = \min\{\tilde{d}_2^j, \sigma_{3n+1}^j\}, \tag{36}$$

where q_{on}^j is the in-flux from the on-ramp. Here we assume that the on-ramp has the absolute merging priority; i.e., the on-ramp demand is satisfied before the mainline freeway demand.

At the lane-drop bottleneck downstream to cell $4n$, we apply the same capacity drop model as (30):

$$v_{4n}^j = \min\{\delta_{4n}^j, (1-\varepsilon_2 \cdot H(\delta_{4n}^j - C))C\}, \tag{37}$$

Note that here we assume that link 5 is always uncongested, and its supply is always C. Thus, when $d_{4n}^j > C$, the maximum discharging flow-rate is dropped to $(1-\varepsilon_2)C$; otherwise, the maximum discharging flow-rate is C.

We can update the two queues as in the following:

$$\lambda_1^{j+1} = \lambda_1^j + (d_1^j - \mu_1^j)\Delta t, \tag{38}$$

$$\lambda_2^{j+1} = \lambda_2^j + (d_2^j - q_{on}^j)\Delta t. \tag{39}$$

Then we calculate the arrival and departure cumulative flows of the whole traffic system as

$$A^j = A^{j-1} + (d_1^j + d_2^j)\Delta t, \tag{40}$$

$$D^j = D^{j-1} + (v_{4n}^j + q_{off}^j)\Delta t, \tag{41}$$

where A^j is the cumulative arrival at $j\Delta t$, and D^j the cumulative departure at $j\Delta t$. Here we set $A^0 = D^0 = 0$. We will simulate the traffic dynamics until the network becomes empty again and assume that the maximum number of time steps is J. Then the total travel time can be calculated from cumulative arrival and departure flows as

$$\tau = \sum_{j=1}^{J} (A^j - D^j)\Delta t. \tag{42}$$

3.3.2 Simulation results

We first set $\bar{d}_2 = 0.05C$. From the analysis in Section 3.2, we expect that a successful control strategy can deactivate the upstream bottleneck without activating the downstream one during the peak period. Thus in this case the VSL control is beneficial, improving the system's performance. The simulation results are shown in Figure 3.4. Note that in the contour plots of densities, values smaller than 1 represent uncongested traffic conditions and congested otherwise. From the figures we can see that, before control, only the upstream bottleneck is activated; after control, the upstream bottleneck becomes uncongested, without activating the downstream one. In this case, the VHTs (vehicle hours traveled) in the system are 730 before control and 523 after control, and the price of myopia in (15) equals -28%. That is, the VSL

control strategy improves the system's performance by 28%. This verifies that the VSL strategy implemented in (27) is effective.
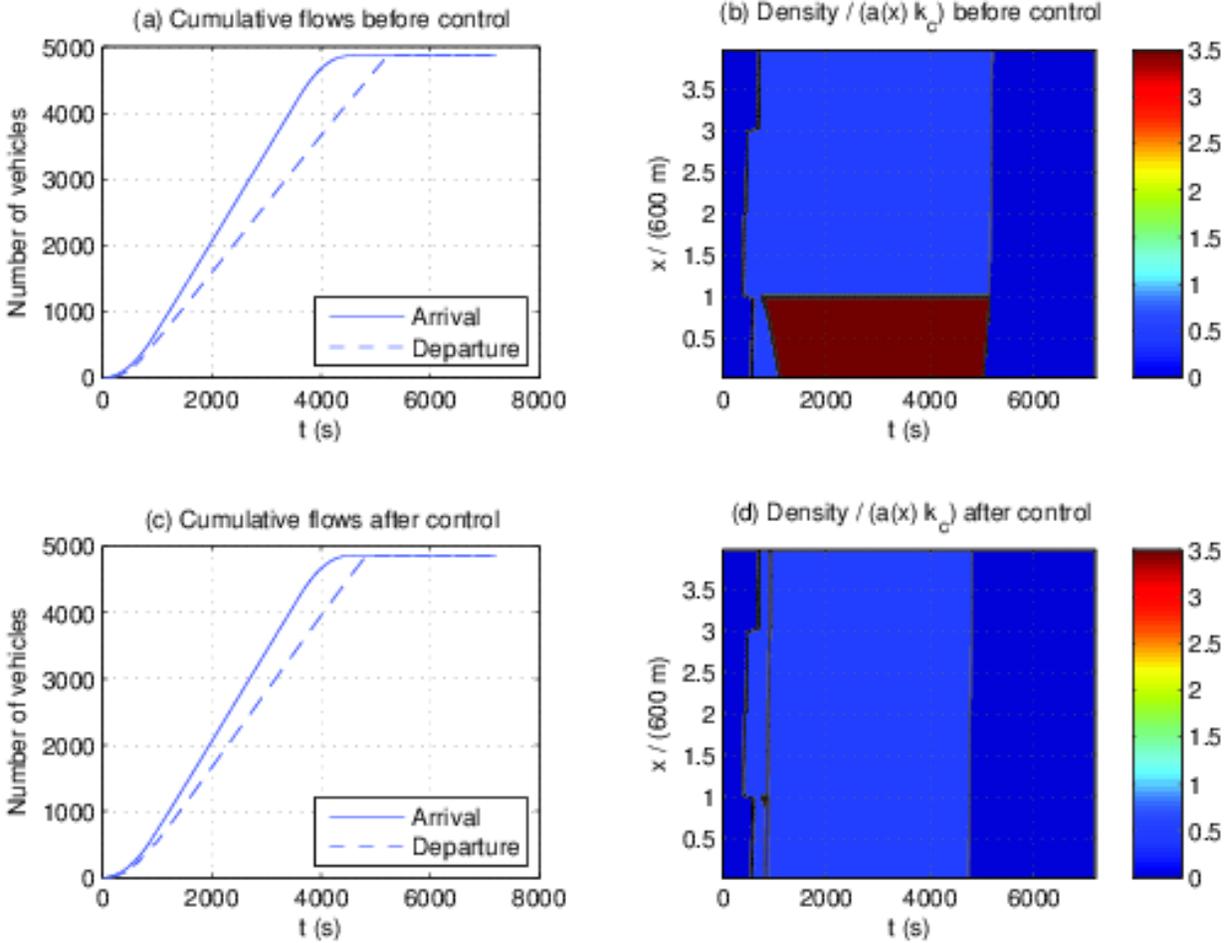

Figure 3.4: Simulation results with $\overline{d}_2 = 0.05C$

We then set $\overline{d}_2 = 0.15C$. From the analysis in Section 3.2, we expect that successful control of the upstream bottleneck can deactivate the upstream bottleneck but activate the downstream one during the peak period, and the paradoxical behavior occurs. The simulation results are shown in Figure 3.5. From the figures we can see that, before control, only the upstream bottleneck is activated. But after control, the upstream bottleneck is first cleared, but the downstream one is activated, and a queue spills back to links 2 and 3. In this case, the VHTs in the system are 733 before control and 792 after control. In this case the price of myopia equals 7.9%; i.e., the VSL control is myopic and increases the total travel time by 7.9%.
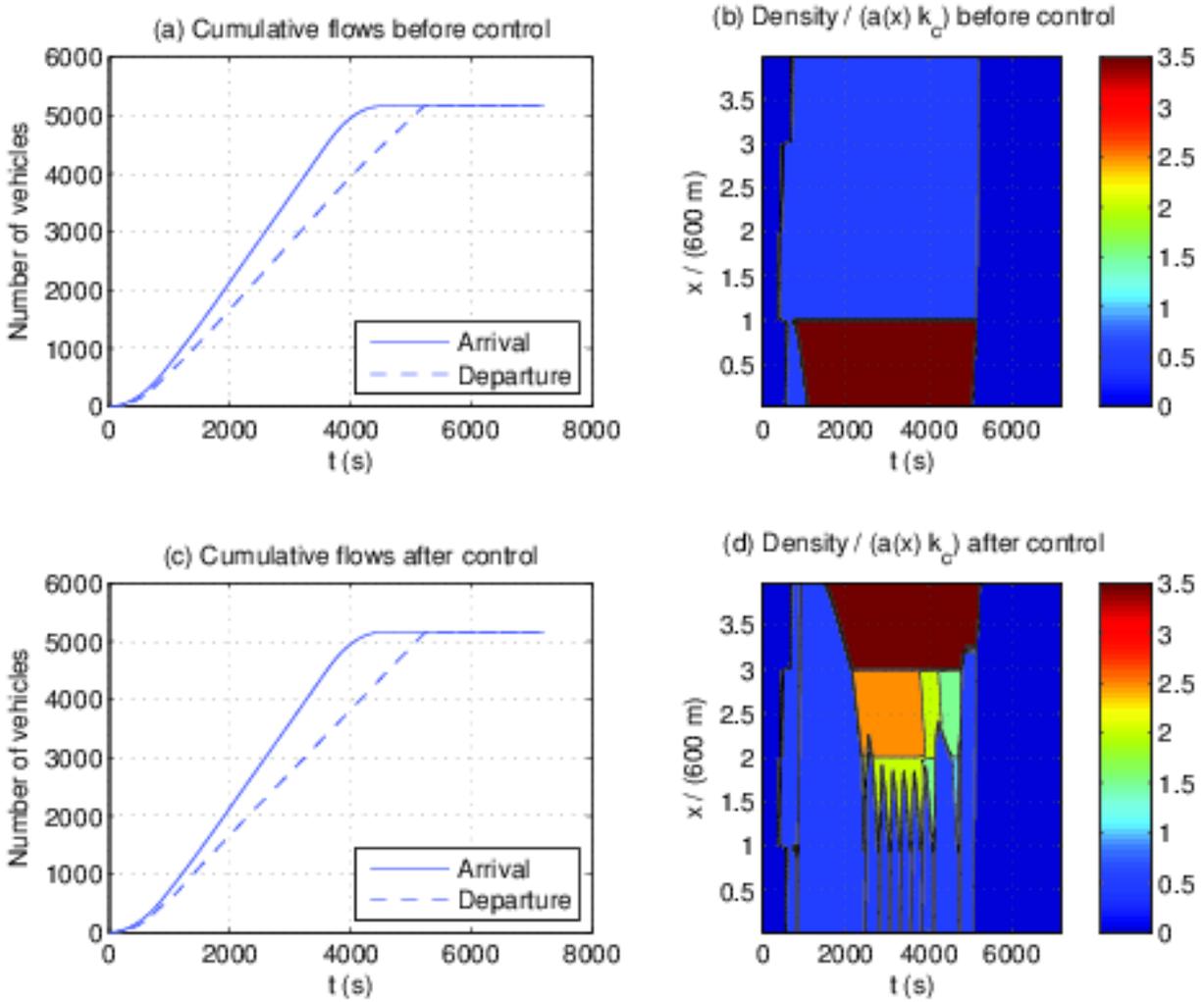
Figure 3.5: Simulation results with $\bar{d}_2 = 0.15C$

We finally set $\bar{d}_2 = 0.5C$. The simulation results are shown in Figure 3.6. From the figures we can see that, before control, both bottlenecks are activated. After control, the upstream bottleneck is partially cleared, but the congestion time of the downstream one is prolonged. In this case the VHTs in the system are 1752 before control and 1787 with control. In this case the price of myopia equals 2%, and the VSL control also leads to worse system performance.
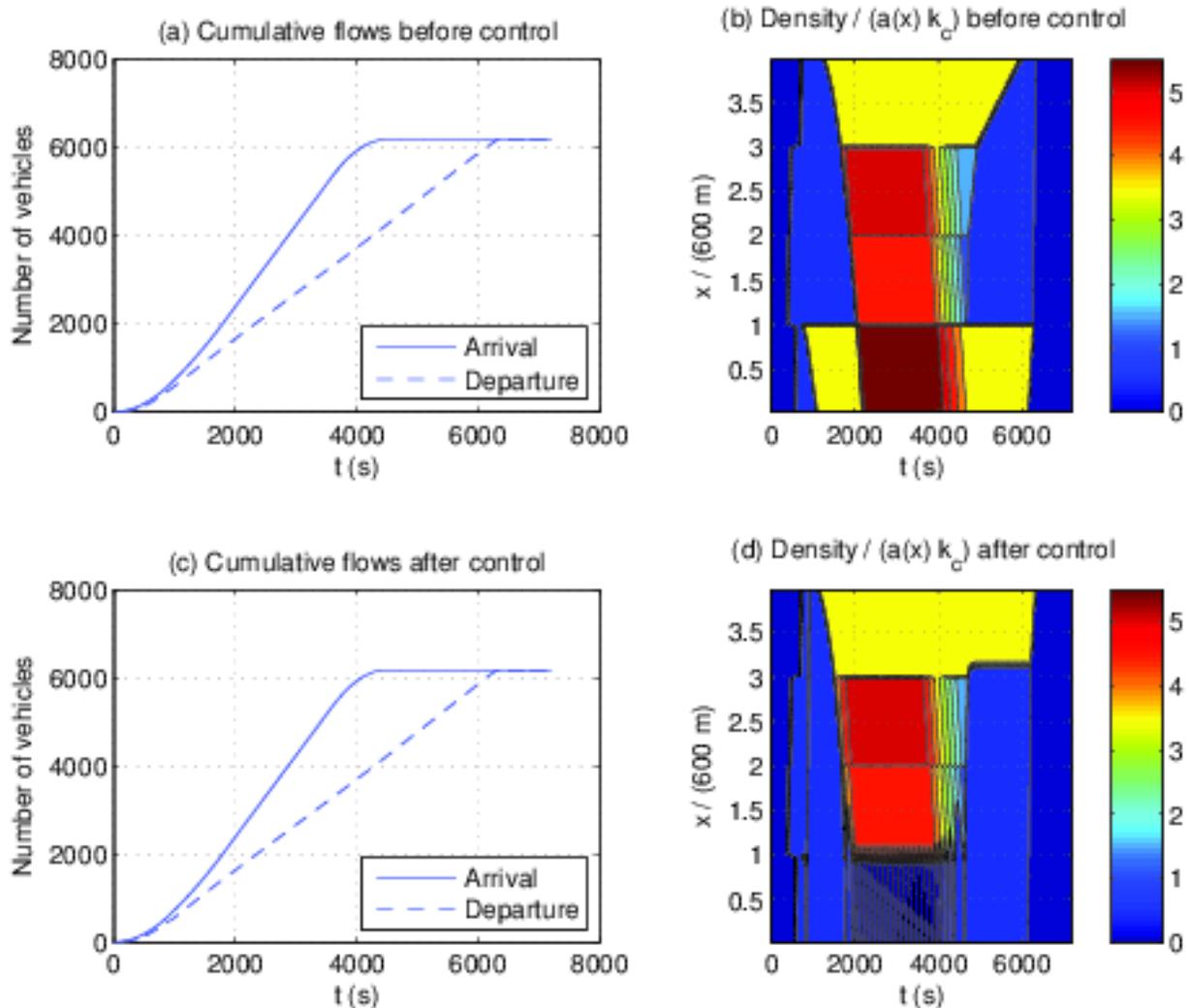
Figure 3.6: Simulation results with $\bar{d}_2 = 0.5C$

In Figure 3.7 , we show the prices of myopia for different on-ramp demands. For each on-ramp demand, we simulate 100 random scenarios and calculate the average price of myopia. From the simulation results we can see that the on-ramp demand plays a critical role in determining the price of myopia: a control strategy is beneficial with low on-ramp demand levels (e.g., $\bar{d}_2 = 0.05C$), and the price of myopia is larger with medium on-ramp demand levels (e.g., $\bar{d}_2 = 0.15C$) than with high on-ramp demand levels (e.g., $\bar{d}_2 = 0.5C$). Such a nonlinear relationship can be explained from traffic dynamics at both bottlenecks: with low demand levels, successful control of the upstream bottleneck deactivates the upstream bottleneck but doesnot activate the downstream bottleneck, and leads to better system performance and, therefore, negative prices of myopia; with medium demand levels, successful control of the upstream bottleneck activates the downstream bottleneck and leads to worse system performance and

therefore large prices of myopia; with high demand levels, both bottlenecks are congested without control, and control of the upstream bottleneck has minimum effect on the system and leads to relatively small prices of myopia.
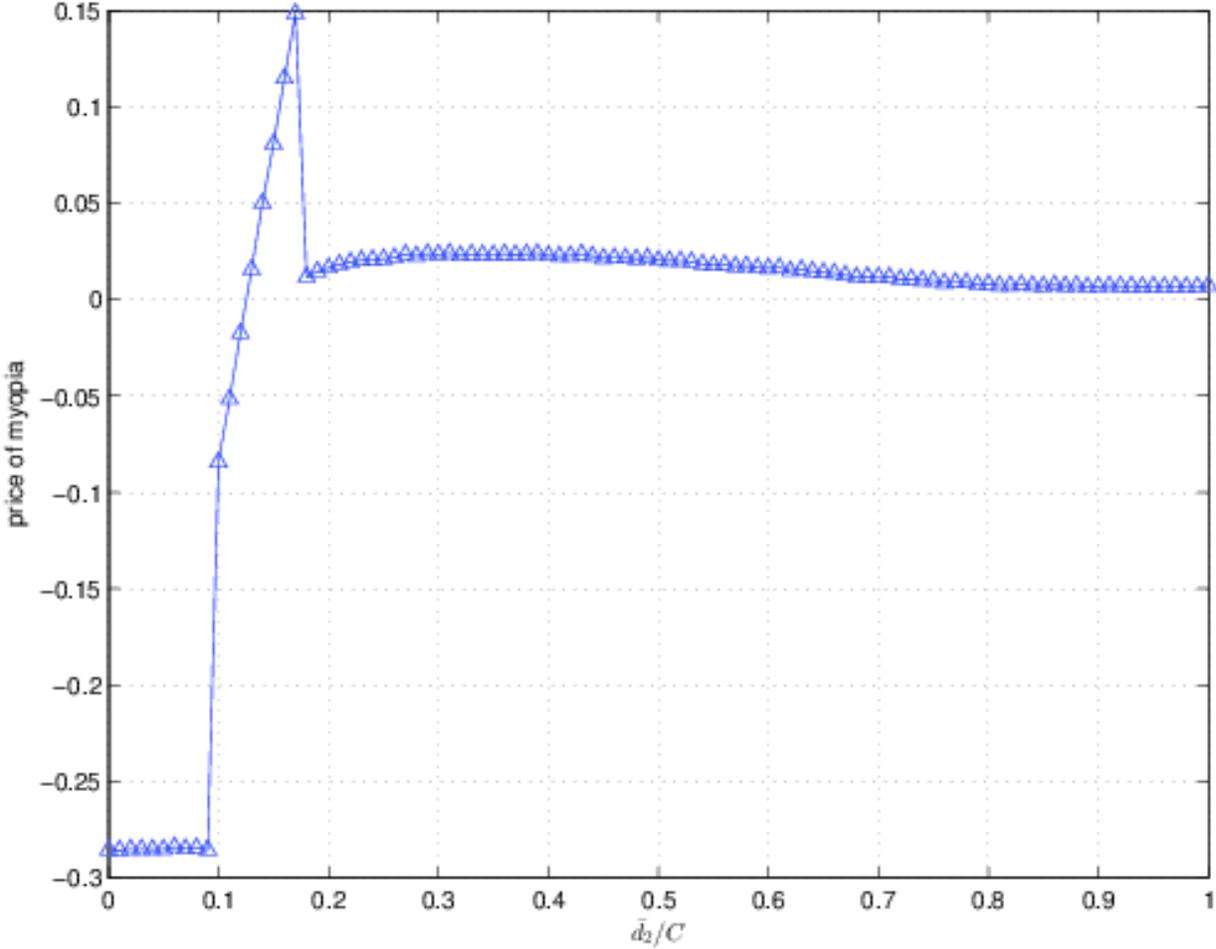

Figure 3.7: Prices of myopia vs on-ramp demands

Chapter 4

Conclusions and Future Research

4.1 Local Ramp Metering

In this report, by combining a simple link queue model to describe traffic dynamics of a merge bottleneck, we were able to show analytically the hysteresis imposed by the capacity drop phenomenon, the reduced reachability region, and the stability range when the merge is controlled by PI-ALINEA.

The reachability is a direct consequence of the hysterisis imposed by the capacity drop phenomenon. The maximum metering rate in which the capacity drop can be avoided is greater than the metering rate necessary to recover from the capacity drop. A quite possible scenario is being possible to avoid the capacity drop, but if a disturbance on the system leads to capacity drop, it might not be possible to recover from it unless the upstream demand ceases. This result is general and regardless of the control strategy.

This is a disadvantage of ramp metering compared to variable speed limit. A reduced speed lower the upstream flow while a lower metering rate reduces the ramp flow. In general, the ramp demand is a small share of the total demand and reducing may not be enough to reduce the total demand. On the other hand, variable speed limit moves the congestion upstream which can hit upstream off-ramps first. Ramp metering moves the congestion from the lane drop to on-ramps, therefore avoiding the congestion

We derived the stability range for the (PI-)ALINEA, one of the most studied ramp metering algorithms. Considering the capacity drop phenomenon, ALINEA can lead the system to the density in which yields maximum throughput if it is in the stability region, that is, theoretically, the target density can be the critical downstream density. In practice, a "slightly undercritical" [Papageorgiou et~al.(1997)Papageorgiou, Hadj-Salem, and Middelham] is set. From the model proposed it is possible to show the reason: the asymmetrical effect of a small disturbance; a small decrease on the upstream demand leads to a small decrease on the out-flux; a small increase, however, can trigger the capacity drop and severely decrease the out-flux. Therefore, a slightly undercritical target occupancy avoids the capacity drop at expense of lower out-flux.

In the future, we will be interested in the ramp metering problem subject to stochastic disturbances in both upstream demands and downstream capacity. In addition, we will be considering the impacts of the on-ramp queue overriding rule as well as practical implementation issues.

4.2 Global Ramp Metering

In this report, we used an example of a freeway control system to demonstrate the occurrence of a paradoxical behavior, in which successful control of an active bottleneck could lead to worse system performance. We first discussed stationary states at a bottleneck with capacity drop and the impacts of local feedback control strategies. Then we derived sufficient conditions in terms of demands, capacity drop magnitudes, and turning ratios for the occurrence of such a paradoxical behavior in stationary states. We further defined the price of myopia, as the increased percentage in the total travel time caused by myopic traffic control strategies. With Cell Transmission Model simulations, we demonstrated that under dynamic and random conditions such a paradoxical behavior still occurs. We found that a local control strategy can be beneficial with low demand levels, and the price of myopia is higher with medium demand levels than with high demand levels.

This study definitely establishes that myopic traffic control strategies can be detrimental to the whole system. It also reveals one underlying mechanism of such a paradox: successful local control activates the downstream, otherwise dormant, bottleneck, and the resulted queue further blocks the upstream bottleneck. Moreover, we demonstrate that the phenomenon occurs under a wide range of conditions in a simple network. One may argue that the model network is quite small, but the occurrence of the paradox in such a small network exactly suggests that it can be more prevalent in a larger network with more complicated interactions. In addition, a new concept, the price of myopia, the price of myopia, was introduced to quantify such negative effects of myopic control strategies. The analytical and simulation methods developed in this study can also be useful to examine the effects of various traffic control strategies for a large road network.

The paradox in this study is analogous to the Braess paradox, where the total travel time could be increased by adding a new road to a network [Braess et~al.(2005)Braess, Nagurney, and Wakolbinger]. Therefore, analysis and quantification of such a paradox can have confound implications for traffic control as that of the Braess paradox for network design. This study clearly shows that coordinated traffic control is not only beneficial, but also necessary, for a complex road network, as otherwise we may have to pay a price of myopia. However, to define the right scope of coordination is a critical question to answer before designing coordinated control strategies. For example, one insight obtained from this study is that one needs to include all bottlenecks that can be potentially activated by the control strategy. Otherwise, we may still have to pay a price of myopia. For a large network, we can start with identifying activate bottlenecks and quantify the price of myopia before and after various control strategies. Once the scope of coordination is determined, one may want to augment the local control law in (27) with global information feedback, e.g., from cell $4n$ on link 4; another approach is to implement control at both bottlenecks. Another insight is that we may be able to improve traffic flow by creating artificial bottlenecks at certain locations. In the future, we will also be interested in

possible paradoxical behaviors in other traffic control systems, including signalized arterial networks and distributed control with connected and automated vehicles.

References

1. [Allaby et~al.(2006)Allaby, Hellinga, and Bullock] Allaby, P., Hellinga, B., Bullock, M., 2006. Variable speed limits: safety and operational impacts of a candidate control strategy for an urban freeway. In: Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE. IEEE, pp. 897–902.

2. Aström, K. J., Murray, R. M., 2010. Feedback systems: an introduction for scientists and engineers. Princeton university press.

3. Banks, J. H., 1991. The two-capacity phenomenon: some theoretical issues. Transportation Research Record: Journal of the Transportation Research Board 1320, 234–241.

4. Bertini, R., Leal, M., 2005. Empirical study of traffic features at a freeway lane drop. Journal of Transportation Engineering 131 (6), 397–407.

5. Braess, D., Nagurney, A., Wakolbinger, T., 2005. On a paradox of traffic planning. Transportation Science 39 (4), 446–450.

6. Carlson, R., Papamichail, I., Papageorgiou, M., Messmer, A., 2010. Optimal motorway traffic flow control involving variable speed limits and ramp metering. Transportation Science 44 (2), 238–253.

7. Carlson, R. C., Papamichail, I., Papageorgiou, M., 2011. Local feedback-based mainstream traffic flow control on motorways using variable speed limits. IEEE Transactions on Intelligent Transportation Systems 12 (4), 1261–1276.

8. Carlson, R. C., Papamichail, I., Papageorgiou, M., 2013. Comparison of local feedback controllers for the mainstream traffic flow on freeways using variable speed limits. Journal of Intelligent Transportation Systems 17 (4), 268–281.

9. Cascetta, E., Punzo, V., Montanino, M., 2011. Empirical analysis of effects of automated section speed enforcement system on traffic flow at freeway bottlenecks. Transportation Research Record: Journal of the Transportation Research Board 2260 (1), 83–93.

10. Cassidy, M. J., Bertini, R. L., 1999. Some traffic features at freeway bottlenecks. Transportation Research Part B: Methodological 33 (1), 25–42.

11. Cassidy, M. J., Rudjanakanoknad, J., 2005. Increasing the capacity of an isolated merge by metering its on-ramp. Transportation Research Part B: Methodological 39 (10), 896–913.

12. Chung, K., Rudjanakanoknad, J., Cassidy, M. J., 2007. Relation between traffic density and capacity drop at three freeway bottlenecks. Transportation Research Part B: Methodological 41 (1), 82–95.

13. Courant, R., Friedrichs, K., Lewy, H., 1928. Über die partiellen differenzengleichungen der mathematischen physik. Mathematische Annalen 100 (1), 32–74.

14. Daganzo, C. F., 1995. The cell transmission model, part ii: network traffic. Transportation Research Part B: Methodological 29 (2), 79–93.

15. Daganzo, C. F., 1999. Remarks on traffic flow modeling and its applications. In: Brilon, W., Huber, F., Schreckenberg, M., Wallentowitz, H. (Eds.), Proceedings of Traffic and Mobility: Simulation, Economics and Environment. Springer Verlag, pp. 105–115.

16. Daganzo, C. F., Gayah, V. V., Gonzales, E. J., 2011. Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability. Transportation Research Part B: Methodological 45 (1), 278–288.
17. Diakaki, C., Papageorgiou, M., McLean, T., 2000. Integrated traffic-responsive urban corridor control strategy in glasgow, scotland: Application and evaluation. Transportation Research Record: Journal of the Transportation Research Board (1727), 101–111.
18. Federal highway administration, 1985. Special Report 209: Highway Capacity Manual. TRB, National Research Council, Washington, D.C.
19. Geroliminis, N., Sun, J., 2011. Hysteresis phenomena of a macroscopic fundamental diagram in freeway networks. Transportation Research Part A: Policy and Practice 45 (9), 966–979.
20. Golob, T. F., Recker, W. W., 2003. Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions. Journal of transportation engineering 129 (4), 342–353.
21. Gomes, G., Horowitz, R., 2003. A study of two onramp metering schemes for congested freeways. In: American Control Conference, 2003. Proceedings of the 2003. Vol. 5. IEEE, pp. 3756–3761.
22. Gomes, G., Horowitz, R., Kurzhanskiy, A. A., Varaiya, P., Kwon, J., 2008. Behavior of the cell transmission model and effectiveness of ramp metering. Transportation Research Part C: Emerging Technologies 16 (4), 485–513.
23. Greenberg, H., Daou, A., 1960. The control of traffic flow to increase the flow. Operations Research 8 (4), 524–532.
24. Haberman, R., 1977. Mathematical models. Prentice Hall, Englewood Cliffs, NJ.
25. Hall, F. L., Agyemang-Duah, K., 1991. Freeway capacity drop and the definition of capacity. Transportation Research Record: Journal of the Transportation Research Board 1320, 91–98.
26. Han, Y., Yuan, Y., Hegyi, A., Hoogendoorn, S. P., 2015. Linear quadratic mpc for integrated route guidance and ramp metering. In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems. IEEE, pp. 1150–1155.
27. Hegyi, A., De Schutter, B., Hellendoorn, H., 2005. Model predictive control for optimal coordination of ramp metering and variable speed limits. Transportation Research Part C 13 (3), 185–209.
28. Jin, H.-Y., Jin, W.-L., 2014. Control of a lane-drop bottleneck through variable speed limits. Transportation Research Part C: Emerging Technologies.
29. Jin, W., Zhang, M., 2001. Evaluation of on-ramp control algorithms. California Partners for Advanced Transit and Highways (PATH).
30. Jin, W.-L., 2010. Continuous kinematic wave models of merging traffic flow. Transportation Research Part B 44 (8-9), 1084–1103.
31. Jin, W.-L., 2012. A link queue model of network traffic flow. arXiv preprint arXiv:1209.2361.
32. Jin, W.-L., 2013. Stability and bifurcation in network traffic flow: A poincaré map approach. Transportation Research Part B: Methodological 57, 191–208.
33. Jin, W.-L., 2015 a . On the existence of stationary states in general road networks. Transportation Research Part B: Methodological 81, 917–929.
34. Jin, W.-L., 2015 b . Point queue models: A unified approach. Transportation Research Part B: Methodological 77, 1–16.

35. Jin, W.-L., Gan, Q.-J., Lebacque, J.-P., 2015. A kinematic wave theory of capacity drop. Transportation Research Part B 81(1), 316–329.

36. Kalman, R. E., et al., 1960. Contributions to the theory of optimal control. Bol. Soc. Mat. Mexicana 5 (2), 102–119.

37. Karafyllis, I., Papageorgiou, M., 2014. Stability results for simple traffic models under pi-regulator control. IMA Journal of Mathematical Control and Information, dnu040.

38. Kim, K., Cassidy, M., 2012. A capacity-increasing mechanism in freeway traffic. Transportation Research Part B 46 (9), 1260–1272.

39. Kotsialos, A., Papageorgiou, M., Mangeas, M., Haj-Salem, H., 2002. Coordinated and integrated control of motorway networks via non-linear optimal control. Transportation Research Part C: Emerging Technologies 10 (1), 65–84.

40. Lebacque, J.-P., 1996. The godunov scheme and what it means for first order traffic flow models. In: Internaional symposium on transportation and traffic theory. pp. 647–677.

41. Levinson, D., Zhang, L., 2006. Ramp meters on trial: Evidence from the twin cities metering holiday. Transportation Research Part A: Policy and Practice 40 (10), 810–828.

42. Liberzon, D., 2012. Switching in systems and control. Springer Science & Business Media.

43. Lighthill, M. J., Whitham, G. B., 1955. On kinematic waves. ii. a theory of traffic flow on long crowded roads. In: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. Vol. 229. The Royal Society, pp. 317–345.

44. Lin, H., Antsaklis, P. J., 2009. Stability and stabilizability of switched linear systems: a survey of recent results. Automatic control, IEEE Transactions on 54 (2), 308–322.

45. Lu, X.-Y., Varaiya, P., Horowitz, R., Su, D., Shladover, S. E., 2010. A new approach for combined freeway variable speed limits and coordinated ramp metering. In: Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on. IEEE, pp. 491–498.

46. Maggi, L., Sacone, S., Siri, S., 2015. Freeway traffic control considering capacity drop phenomena: comparison of different mpc schemes. In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems. IEEE, pp. 457–462.

47. Munjal, P. K., Hsu, Y. S., Lawrence, R. L., 1971. Analysis and validation of lane-drop effects of multilane freeways. Transportation Research 5 (4), 257–266.

48. Newell, G. F., 1993. A simplified theory of kinematic waves in highway traffic, part i: General theory. Transportation Research Part B: Methodological 27 (4), 281–287.

49. Normey-Rico, J. E., Camacho, E. F., 2008. Dead-time compensators: A survey. Control engineering practice 16 (4), 407–428.

50. Papageorgiou, M., 1998. Some remarks on macroscopic traffic flow modelling. Transportation Research Part A: Policy and Practice 32 (5), 323–329.

51. Papageorgiou, M., H. Hadj-Salem, J.-M. B., 1991. Alinea: A local feedback control law for on-ramp metering. Transportation Research, 58–64.

52. Papageorgiou, M., Hadj-Salem, H., Middelham, F., 1997. Alinea local ramp metering: Summary of field results. Transportation Research Record: Journal of the Transportation Research Board 1603 (1), 90–98.

53. Papageorgiou, M., Kotsialos, A., 2000. Freeway ramp metering: An overview. In: Intelligent Transportation Systems, 2000. Proceedings. 2000 IEEE. IEEE, pp. 228–239.

54. Papamichail, I., Kotsialos, A., Margonis, I., Papageorgiou, M., 2010. Coordinated ramp metering for freeway networks–a model-predictive hierarchical control approach. Transportation Research Part C: Emerging Technologies 18 (3), 311–331.

55. Persaud, B., Yagar, S., Brownlee, R., 1998. Exploration of the breakdown phenomenon in freeway traffic. Transportation Research Record: Journal of the Transportation Research Board 1634, 64–69.

56. Richards, P. I., 1956. Shock waves on the highway. Operations research 4 (1), 42–51.

57. Roughgarden, T., 2005. Selfish Routing And The Price Of Anarchy. MIT Press.

58. Schrank, D., Eisele, B., Lomax, T., 2012. The 2012 urban mobility report. Texas A&M Transportation Institute. The Texas A&M University System.

59. Smaragdis, E., Papageorgiou, M., Kosmatopoulos, E., 2004. A flow-maximizing adaptive local ramp metering strategy. Transportation Research Part B: Methodological 38 (3), 251–270.

60. Sun, X., Horowitz, R., 2005. A localized switching ramp-metering controller with a queue length regulator for congested freeways. In: American Control Conference, 2005. Proceedings of the 2005. IEEE, pp. 2141–2146.

61. Sun, X., Horowitz, R., 2006. Set of new traffic-responsive ramp-metering algorithms and microscopic simulation results. Transportation Research Record: Journal of the Transportation Research Board (1959), 9–18.

62. Sun, Z., Ge, S. S., Lee, T. H., 2002. Controllability and reachability criteria for switched linear systems. Automatica 38 (5), 775–786.

63. Wang, Y., Kosmatopoulos, E. B., Papageorgiou, M., Papamichail, I., 2014. Local ramp metering in the presence of a distant downstream bottleneck: Theoretical analysis and simulation study. Intelligent Transportation Systems, IEEE Transactions on 15 (5), 2024–2039.

64. Wang, Y., Papageorgiou, M., Gaffney, J., Papamichail, I., Rose, G., Young, W., 2010. Local ramp metering in random-location bottlenecks downstream of metered on-ramp. Transportation Research Record: Journal of the Transportation Research Board (2178), 90–100.

65. Wiggins, S., 2003. Introduction to applied nonlinear dynamical systems and chaos. Vol. 2. Springer Science & Business Media.

66. Yang, H., Gan, Q., Jin, W.-L., 2011. Calibration of a family of car-following models with retarded linear regression methods. Proceedings of Transportation Research Board Annual Meeting.

67. Yang, H., Jin, W.-L., 2014. A control theoretic formulation ofgreen driving strategies based on inter-vehicle communications. Transportation Research Part C 41, 48–60.

68. Zhang, J., Chang, H., Ioannou, P. A., 2006. A simple roadway control system for freeway traffic. In: American Control Conference. IEEE.