

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

The aggregate effect of turns on urban traffic networks

Permalink

<https://escholarship.org/uc/item/4rr4g9wm>

Author

Gayah, Vikash Varun

Publication Date

2012

Peer reviewed|Thesis/dissertation

The aggregate effect of turns on urban traffic networks

by

Vikash Varun Gayah

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy
in

Engineering – Civil and Environmental Engineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Carlos F. Daganzo, Chair

Professor Michael J. Cassidy

Professor Robert Cervero

Professor Samer Madanat

Spring 2012

The aggregate effect of turns on urban traffic networks

Copyright © 2012

by

Vikash Varun Gayah

Abstract

The aggregate effect of turns on urban traffic networks

by

Vikash Varun Gayah

Doctor of Philosophy in Engineering – Civil and Environmental Engineering

University of California, Berkeley

Professor Carlos F. Daganzo, Chair

This dissertation furthers the understanding of macroscopic traffic behavior on urban networks. In particular, it focuses on the aggregate effect of turning maneuvers in networks with multiple routes. The presence of turning between different routes on a network is found to have two effects: 1) it causes chaotic and inefficient behavior, especially when the network is congested; and, 2) it may reduce maximum vehicle flows across the network. Fortunately, this work finds that designing networks that are redundant (i.e., have multiple unique routes between all origin-destination pairs), helping drivers avoid locally congested regions and limiting the rate at which vehicles are allowed to enter the network can help mitigate the first effect, and this allows the network to operate more efficiently. It is also found that the second effect may not always be harmful—lower network flows do not necessarily result in decreased network efficiency if the lower flows are accompanied by more direct routing between origins and destinations. In fact, two-way networks, which accommodate conflicting left-turns and result in lower maximum vehicle flows than one-way networks, are found to serve trips at a higher rate because drivers travel shorter distances on average. Thus, in many cities, maximum network efficiency can be improved by converting one-way streets to two-way operation.

To my family, for their love, support and sacrifice,
and to Barbara Pennington, for helping me to believe in myself.

Contents

Contents	ii
List of Figures	iv
Acknowledgments	vi
1 Introduction	1
1.1 Background on aggregate models of urban traffic	2
1.2 Research questions	7
1.3 Main contributions	7
1.4 Organization	8
2 Simulation evidence of unstable behavior	10
2.1 Basic grid network	10
2.2 A two-ring idealization	11
2.3 Interactive simulation	12
2.4 Phenomena to observe	15
2.5 Major findings	18
3 Analytical explanation of instability	20
3.1 A two-bin idealization	20
3.2 Analysis of a homogeneous square grid	23
3.3 Effect of network heterogeneity	29
3.4 Effect of driver adaptation	33
3.5 Major findings and implications for more realistic networks	35
4 Macroscopic network dynamics during rush hour	39
4.1 Generalization of two-bin system	39
4.2 System dynamics	40
4.3 Patterns in the flow-density relationship	47
4.4 Effect of adaptive drivers	50
4.5 Major findings and implications for more realistic networks	57

5	Effect of turning on network efficiency	59
5.1	Capacity formula	60
5.2	Networks considered	62
5.3	Network parameters	65
5.4	Network comparison	69
5.5	Extension to more realistic networks	74
5.6	Major findings and discussion	75
6	Conclusions	77
6.1	Summary of major findings	77
6.2	Future work	78
	Bibliography	80

List of Figures

1.1	Comparison of MFDs	5
1.2	Empirical flow-density relationship observed in Toulouse, France . . .	6
2.1	Simple grid network	11
2.2	Triangular fundamental diagram	11
2.3	Two-ring system	12
2.4	Simulation interface	14
2.5	Two-ring behavior with no turning	16
2.6	Two-ring behavior with turning	17
2.7	Two-ring behavior with signals	19
3.1	Intersection control simplification	21
3.2	Two-bin model	21
3.3	State space and resulting flow-density relationship	24
3.4	Possible regimes for symmetric two-bin system	25
3.5	Results of symmetric two-bin system	26
3.6	Jam patterns in a homogeneous network	28
3.7	Flow-density relationship for grid network	29
3.8	Heterogeneous fundamental diagrams	30
3.9	Results of heterogeneous two-bin system	31
3.10	Results of rectangular two-bin system	32
3.11	Results of two-bin system with asymmetric demand	34
3.12	Change in $F_i(k)$ with adaptive drivers	35
3.13	Results of symmetric two-bin system with adaptive drivers	36
3.14	Comparison of MFDs	37
4.1	Neighborhoods within a city network	40
4.2	Two-bin idealization with inflows and outflows	40
4.3	Converging and diverging areas	42
4.4	Loading paths for balanced, exogenous loading	43
4.5	Recovery paths	46
4.6	Patterns in the flow-density plane during an LR cycle	48
4.7	Loading on a more realistic network	51
4.8	Recovery on a more realistic network	52

4.9	Converging and diverging areas with adaptive drivers	53
4.10	Loading and recovery paths with adaptive drivers	55
4.11	Phase paths with random turns and adaptive drivers	56
4.12	Empirical MFDs	58
5.1	Intersection configuration for idealized network	61
5.2	Intersection configurations for different networks	63
5.3	Capacities of two-lane networks	70
5.4	Capacities of three-lane networks	71
5.5	Capacities of four-lane networks	72
5.6	Comparison of one-way and two-way network with left-turn pockets .	73

Acknowledgments

A wise man once said, “turns out not where but who you’re with that really matters.” This has definitely been the case during my time at Berkeley. The people with which I have interacted, shared ideas, collaborated, laughed, stressed, played basketball and grown up have changed me in ways that I will be forever grateful. I would like to acknowledge the following people in particular.

First and foremost, I would like to thank Carlos Daganzo. As my research advisor, Carlos has helped guide the direction of my dissertation work, pushed me to explore new ideas and shown me the true meaning of academic research. For this I will be forever grateful. He has taught me so much in our short time together (both in the classroom and in the office), and I hope these lessons continue well on into the future. I will always be amazed by his ability to whittle a problem down into its most basic component and derive meaningful insights from analyzing something simple. This is a skill I hope to develop myself, though I will never be able to master it as he has. I would also like to thank him for his friendship and the invaluable advice he has given me about my career and personal life. It has truly been a joy working with him the past five years.

I would also like to thank the rest of the faculty at UC Berkeley for their support and guidance. In particular, I would like to acknowledge Michael Cassidy, who served on my dissertation committee and who I also consider as an advisor and mentor. I appreciate the time he spent with me discussing research ideas and career issues, even though I was never formally his student. He has also taught me how to communicate effectively in the classroom, and I hope I can one day be as good of a teacher as he is. I would also like to thank the remaining members of my dissertation committee, Samer Madanat and Robert Cervero. The feedback they provided during my qualifying exam and dissertation workshop has been invaluable to this dissertation. Special thanks also goes to Alexander Skabardonis for his ability to brighten up any day. Life in 416 McLaughlin Hall would not have been as manageable without him.

I have been lucky to interact with so many amazing students during my time at UC Berkeley. I did not know it before I accepted admission, but one of the biggest strengths of the transportation engineering program is quality and character of the students in the program. I would like to acknowledge all of the members of 416 McLaughlin Hall who have made the office a home away from home. In particular, I would like to thank my officemates: Celeste Chavis, Eleni Christofa, Weihua Gu and Karthik Sivakumaran. I am grateful for their friendships and I will always cherish the time we spent together in 416G. They are especially deserving of my gratitude since they have put up with my singing, humming, drumming, air drumming, air violining and dancing in the office, without forcing me to change desks. I couldn’t have asked for a better group to see each and every day. Thanks also to Nikolas Gerolimis and Offer Grembek for their friendship and assistance, especially during my first semester at Berkeley. Their willingness to chat and helpfulness in every way made me realize that I made the right choice by moving to Berkeley. I would also like to thank Eric Gonzales for serving as a sounding board for ideas, a colleague to collaborate with, a

senior student to get advice from, a friend to grab a beer with, and much more. I am also grateful to many colleagues and friends outside of 416 that have help to make this PhD experience enjoyable. Special thanks to Daniel Hennessey, Gurkaran Buxi, DJ Gaker, Aditya Medury, Juan Argote and Michael Seelhorst. Additionally, thanks to the ITS Library staff, including Kendra Levine, Chris Coad and Rita Evans, for their assistance in the ITS library whenever I needed it.

I would also like to thank the most important person in Berkeley to me, Ilgin Guler, for everything. She has truly been my constant for the last two and a half years. As my best friend, she has provided me with the support, love, encouragement and (when necessary) kick-in-the-pants needed to get through this. This process has been very painful at times, but I would gladly do it again knowing that she would be there with me.

Though not in Berkeley, I would also like to thank my family—my parents, Vishram and Barbara, and especially my sister, Adeta. My parents gave up their lives in Trinidad and moved to the United States in order to give my sister and I better opportunities in life. Not a day goes by when I am not grateful for this. I strive everyday to take advantage of these opportunities and hope that in doing so, I can honor them and their sacrifice. My sister inspires me everyday with her work ethic and uplifting attitude. Life is so much more bearable knowing that she is there to talk to when things get rough. Without their love and support, I would never have been able to start, much less complete, this daunting task. I hope I have made them proud.

This research was supported by the following sources: the University of California Transportation Center; the University of California, Berkeley Volvo Center for Future and Urban Transport; the Federal Highway Administration Dwight D. Eisenhower Fellowship program; and, the National Science Foundation.

Chapter 1

Introduction

Traffic congestion is a problem that continues to plague major urban areas, both in the United States and around the world. According to one estimate, drivers in urban areas of the United States waste an average of 36 hours and 24 gallons of fuel annually due to delay caused by congestion, and this number continues to increase every year (Schrank & Lomax, 2009). In the past, planners and engineers have combated this congestion by building new roadways to accommodate increasing car demand. However, constantly expanding the roadway infrastructure is not a feasible option for several reasons. The most intuitive reason is that this new infrastructure is expensive. Furthermore, many urban areas are completely ‘built out’ such that there is no space to accommodate new roads. Even in areas that are not built out and for which cost is not prohibitively high, building new roads might not be a long-term solution. Researchers have shown that increasing network capacity by building new roads simply attracts more users, a phenomenon known as induced demand (Lee *et al.*, 1999; Cervero, 2002). If this effect is strong, continually expanding street networks may simply exacerbate congestion.

Another strategy to combat urban congestion might be to implement control strategies that try to use the existing street infrastructure more efficiently. Some examples of Intelligent Transportation Systems (ITS) technologies used to improve local throughput include traffic signal coordination on arterials, variable message signs to provide information to drivers and advanced incident management strategies. Network-wide (or macroscopic) control strategies are also used around the world to increase the efficiency of urban street networks. For example, London, Rome, Stockholm and Singapore have all implemented congestion pricing strategies that charge vehicles to enter the downtown area (Phang & Toh, 2004; Eliasson & Mattsson, 2006; Litman, 2011). By controlling the number of vehicles in the most congested part of the network (through pricing) higher flows and speeds are achieved and maintained. Zurich has also implemented an advanced traffic signal control scheme that restricts capacity to areas in a network that are congested and instead directs flow to underutilized portions of the network (Nash & Sylvia, 2001).

To rationalize the implementation of these macroscopic control strategies and optimize their effects (e.g., to determine appropriate pricing schemes or the maximum

amount of traffic that should be allowed in the downtown at once), planners and engineers need to accurately predict the impact of these strategies on traffic flow. Microscopic traffic simulations are typically used to predict the effects of these network-wide strategies by testing different schemes and comparing the simulated results on traffic flow. However, accurate microscopic models take a long time to run and require tremendous amounts of data, including detailed origin-destination tables about travel demand in an urban area that simply might not exist. Additionally, microscopic predictions are highly sensitive to small changes in input data; if inputs are not accurate, the results obtained may not be useful. Lastly, the stochastic nature of these simulations make it difficult to separate background noise in the output from meaningful relationships of interest. For these reasons, there has been a recent movement to model traffic networks macroscopically by using relationships between aggregate traffic variables measured across the entire network. These aggregate measures typically require much less data than detailed models, are more robust to stochastic variations and can even be more accurate than detailed simulations (Daganzo *et al.*, 2012).

However, current understanding of the aggregate behavior of vehicles on urban traffic networks is limited. While robust and reproducible relationships are observed for some networks, other networks have unpredictable behavior that researchers simply cannot explain. Additionally, the relationships that are actually observed (either empirically or through simulation) differ greatly from what would be expected if congestion was uniformly distributed across the network. Also, empirical data show that networks behave differently during the beginning of a rush hour than during the end of a rush hour. With these unexplained phenomena in mind, this dissertation aims to more deeply explore aggregate traffic behavior on urban networks to gain a more fundamental understanding of why the phenomena occur, when they should be expected and their implications for urban traffic management.

Before delving into the research, this chapter presents some preliminary ideas. Section 1.1 furnishes background on aggregate models of urban traffic. Section 1.2 describes the research questions that will be answered as a part of this dissertation. Section 1.3 presents the main contributions of this work. Finally, Section 1.4 provides the organization of the rest of this dissertation.

1.1 Background on aggregate models of urban traffic

Much attention has been focused on modeling traffic behavior at the the individual link level, both microscopically and macroscopically. Over the past forty years, work has also been directed at modeling the aggregate behavior of vehicles across an entire city network. Section 1.1.1 briefly examines some of the earlier work that has been done in the latter vein and explains their shortcomings. Section 1.1.2 then describes a more recent urban traffic model that captures network dynamics, and some evidence of chaotic and unpredictable behavior.

1.1.1 Early examples of macroscopic traffic relationships

The earliest works on macroscopic network behavior examined relationships between urban traffic variables, and can be divided into two basic categories. Models in the first category proposed relationships between average vehicle speed and average network flow (or network production) that were monotonically decreasing (Smeed & Wardrop, 1964; Smeed, 1966; Thomson, 1967; Wardrop, 1968; Zahavi, 1972). The first example of this, Smeed & Wardrop (1964), used data from Central London and proposed that average network flow decreases linearly with the cube of average speed. This relationship was then used with dimensional analysis in Smeed (1966) to derive a function relating the maximum flow that can enter a city area and variables describing the network, such as the fraction of the city area covered by roads and capacity per roadway width. Thomson (1967), Wardrop (1968) and Zahavi (1972) proposed even more complex models and fit these to data, but all of these models predict that average vehicle speed decreases as the average network flow increases. While these models might accurately describe traffic states when the network is operating at or near free flow conditions, they fail to recognize states that arise when the network is congested and approaching gridlock (i.e., states in which both speeds and flow are low). Therefore, more comprehensive models are needed to accurately describe traffic behavior in a network for all possible states that may arise.

Models in the second category assumed a unimodal relationship between average vehicle speed and average network flow (Godfrey, 1969; Herman & Prigogine, 1979; Ardekani & Herman, 1987; Mahmassani *et al.*, 1987; Mahmassani & Peeta, 1993; Olaszewski *et al.*, 1995); these models were more realistic because they could capture both free-flow and congested conditions within the network. The first study to make this realization was Godfrey (1969). That study also showed that a unimodal relationship between average speed and flow in a network would imply a unimodal relationship between average network flow and average network density. Consistent and reproducible relationships between average network flow and density have come to be known as the ‘Macroscopic Fundamental Diagram’ (or MFD). More refined speed-flow models were later created and fit to empirical data. For example, Herman & Prigogine (1979) proposed a two-fluid model of network traffic that established a relationship between average vehicle speed and the fraction of vehicles in a network that are stopped. Using further assumptions, such as assuming the existence of a relationship between the fraction of vehicles stopped and average network density (known as the ergodic assumption), the authors created a physically realistic speed-density relationship of urban traffic networks. Various studies have been done to calibrate and/or modify the two-fluid model. One example, Ardekani & Herman (1987), modified the relationship between the fraction of vehicles stopped and network density, and calibrated the model using data from aerial photographs and ground observations. Other studies, like Mahmassani *et al.* (1987), simulated idealized networks to calibrate macroscopic speed-density relationships. While these models were an improvement over models in the first category of studies, they were not without faults. For one, due to the scarcity of empirical data (especially data for congested network conditions) researchers could

not prove that these relationships were always reproducible. Additionally, the models relied on some key assumptions that were not empirically verified; e.g., the assumed relationship between the fraction of stopped vehicles and average network density in the two-fluid model. Additionally, these models focused solely on equilibrium conditions so they could not be used to study the dynamics of city traffic as it evolves during a rush hour period.

1.1.2 Macroscopic traffic dynamics and evidence of network instability

Daganzo (2007) reintroduced the idea of a reproducible relationship between macroscopic traffic variables on an urban network as a part of an urban traffic dynamics model. That study showed how a city could be modeled as a single reservoir without the need for detailed origin-destination data that are tedious (and sometimes impossible) to obtain. Using this single reservoir model, Daganzo (2007) revealed insights on the gridlock phenomenon that plagues many urban networks during peak periods. The study also argued that, even if the ergodic assumption of the two-fluid model does not hold, a well-defined relationship between average network flow and density (i.e., the MFD) should arise if a network is uniformly loaded so that all links are similarly congested.

Daganzo & Geroliminis (2008) showed that an MFD must arise in single-route networks that have no turning traffic and are uniformly loaded, if vehicles obey the kinematic wave theory of traffic flow (Lighthill & Whitham, 1955). The work also gave explicit formulas for the MFD as a function of the lengths of the links, their individual fundamental diagrams and the intersection control strategies. Further, the study proposed that these formulas would be accurate for homogeneous, redundant multi-route networks with slow-changing demand because network redundancy would allow drivers to switch (or turn) between routes and distribute themselves uniformly across the network. The authors also conjectured that small amounts of random turning would not significantly affect the distribution of traffic across the network.

The existence of the MFD was not experimentally proven in the field until advances in ITS technologies facilitated the collection of requisite urban-scale data. Using simulations of San Francisco, California and empirical data from from Yokohama, Japan, Geroliminis & Daganzo (2008) were the first to provide evidence strongly suggesting that, at least in some instances, an MFD appears to exist for real networks. Yet, although the networks examined in this study approximately meet the conditions outlined in Daganzo & Geroliminis (2008), that reference showed that the analytical prediction overestimated the observed network flows of San Francisco for very high network densities. This phenomenon was also noted in a more recent study that derived the MFD of Nairobi, Kenya using simulation (Gonzales *et al.*, 2009). The authors of that study attributed these lower flows to a lack of network redundancy which prevents drivers from switching from congested links to nearby underutilized links.

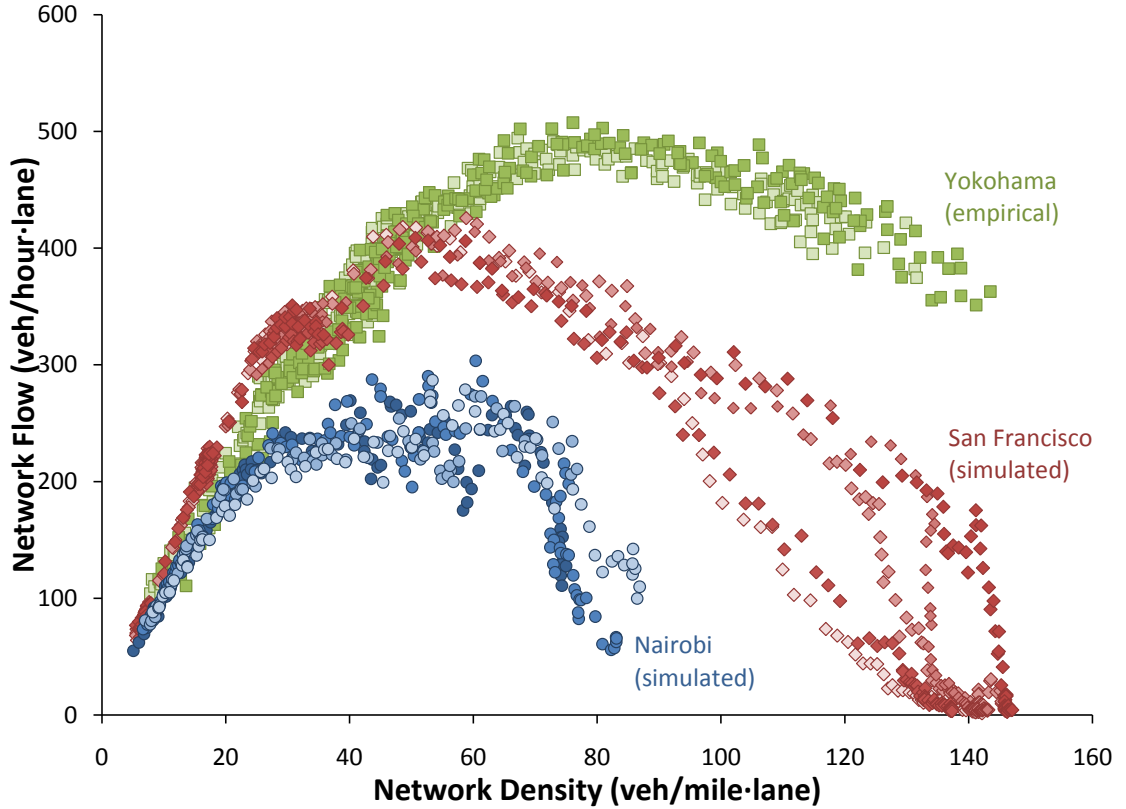


Figure 1.1. Comparison of the macroscopic fundamental diagrams of Yokohama (measured, Geroliminis & Daganzo, 2008), San Francisco (simulated, Geroliminis & Daganzo, 2007) and Nairobi (simulated, Gonzales *et al.*, 2009).

However, reasons for the over-prediction of the MFD of San Francisco’s redundant network were not clear.

In addition to over-prediction of network flows, large scatter is sometimes observed in the MFD for higher network densities. Figure 1.1 shows three MFDs that were reported in Gonzales *et al.* (2009). In this figure, the flow-density relationships for Nairobi, Kenya and San Francisco, California (the two MFDs obtained from simulated data) are well defined for low densities and very scattered for higher densities ($k > 40$ veh/mile-lane for Nairobi, $k > 90$ veh/mile-lane for San Francisco). The study again attributed some of the scatter observed in the Nairobi MFD to the lack of network redundancy, but reasons for the scatter in the San Francisco MFD were not known.

Scatter in the MFD was also noted in simulations of an idealized, completely symmetric, network (Mazloumian *et al.*, 2010). That study suggested that the granular nature of individual vehicles could help to explain some of the scatter that was observed. It also showed that observed network flows change with the spatial variability of vehicle densities across the network—the higher this variability, the lower the ob-

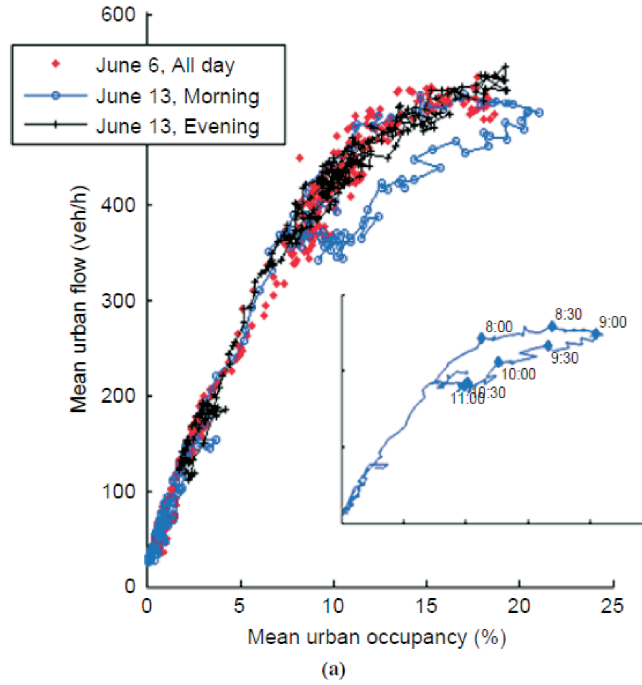


Figure 1.2. Empirical flow-density relationship observed in Toulouse, France (source: Buisson & Ladier, 2009).

served flows. The study also found that observed flows for a given density decreased with time in this particular network; this suggests that spatial variability of link densities might naturally increase with time. However, the reason for this phenomenon was not fully explored. Geroliminis & Sun (2011b) also found the spatial variability of vehicle densities to be a key component in the determination of aggregate network flows. These authors suggested that a well-defined MFD with little scatter and the highest network flows should arise when this spatial variability is low, and that average network flows will be lower when this variability is high.

Empirical data taken from Toulouse, France (Buisson & Ladier, 2009) also shows another phenomenon observed in aggregate flow-density relationships that could not be explained at the time. Figure 1.2 presents the flow-density relationship observed on the urban street network for many rush hour periods. Notice that for most days the data all fall along the same curve, with only a single value of flow observed for each density. However, the observations for June 13, 2008, which are magnified with time stamps on the inset of Figure 1.2, exhibit multiple flow values for some densities. In fact, there appears to be a clear trend with higher flows as the density increases than as it decreases. This shows that networks may behave differently during the beginning of a rush hour (the onset of congestion) than at the end of a rush hour (the dissipation of congestion). The reason for this behavior, and if it should be expected on other networks, is not currently known.

1.2 Research questions

Based on the literature, it is clear that the current understanding of aggregate vehicle behavior on urban networks is limited: the issues of scatter in the MFD, theoretical over-prediction of the MFD and asymmetric network behavior during a rush hour still remain. While previous work suggested that a lack of network redundancy or vehicular granularity might cause this behavior, there might be another, more fundamental, reason for this—drivers switching (or turning) between multiple routes in the network. The formulas developed by Daganzo & Geroliminis (2008) are for a network with a single route (i.e., no turning), and it was proposed that they should be accurate when vehicles are allowed to turn between routes. However, the data shown in Figure 1.1 may indicate that multiple steady-state flows arise on multi-route networks for high densities—in some cases the scatter is as large as 50% and this is hard to explain by randomness alone. If networks have multiple equilibrium states and each of these is associated with a different amount of spatial variability, this could explain the high level of scatter observed in these relationships. This is not an outlandish conjecture since multiple steady states have been proven to exist for very simple networks in which vehicles can choose between different routes (Daganzo, 1998; Jin, 2009). Also, if turning causes traffic to naturally tend toward uneven spatial distributions with time, this might explain why lower flows are observed as the network recovers from a rush hour period.

Therefore, the questions that this dissertation seeks to answer are:

- Does the presence of vehicles that can turn between different routes cause multiple equilibrium states to emerge in urban networks?
- Can these multiple equilibrium states explain the chaotic behavior that is observed in aggregate traffic relationships for these networks?
- If multiple equilibrium states exist due to turning, does this affect network dynamics?
- Does turning also affect other properties of an urban traffic network, such as its maximum efficiency?

1.3 Main contributions

This dissertation produces both scientific and practical benefits. The main contributions of this research are:

- The creation of an interactive simulation that systematically demonstrates the chaotic behavior that arises on a closed multi-route network when vehicles are

allowed to turn between unique routes. This simulation can be used as a teaching tool to demonstrate the phenomena caused by turning between different routes on steady-state behavior.¹

- The development of an analytical model that proves: the existence of multiple equilibrium states in congested networks, bifurcations in the set of stable equilibrium states, a tendency of vehicles to move towards uneven distributions in congested networks, lower flows due to these uneven distributions, and multi-valuedness in the plot of aggregate flow vs. density. This model also shows that even vehicle distributions are not always sustainable when drivers follow rigid routes; in this case, current macroscopic theory will over-predict expected flows on a network.
- Using the analytical model, a demonstration that the instability present in urban networks can be mitigated by the presence of drivers that adapt their routes in response to congestion. This highlights the importance of designing networks that are redundant and well-connected, and providing drivers with real-time information about network conditions so that they can avoid congested areas.
- The development of an analytical model that describes how urban network dynamics differ during the beginning of a rush hour as compared to the end of a rush hour, and the effect of this asymmetric behavior on patterns that are observed in macroscopic traffic relationships. This model also shows that the presence of adaptive drivers can reduce this asymmetry.
- The creation of an interactive simulation that demonstrates the asymmetric network dynamics caused by turning between multiple routes in an open traffic network. This simulation can also be used as a teaching tool to demonstrate the phenomena caused by between route turning on network dynamics.²
- The development of an analytical methodology that describes both the vehicle-moving and trip-serving capacities of homogeneous networks based on the treatment of conflicting left-turning maneuvers between different routes. This model can provide valuable insights to city planners and traffic engineers about the operation of urban traffic networks.

1.4 Organization

The remainder of this dissertation is organized as follows. Chapter 2 introduces the grid network that will be considered throughout this text, and describes one of the

¹This simulation is available for use online at http://www.ce.berkeley.edu/~daganzo/Simulations/two_ring_sim.html.

²This simulation is available for use online at <http://www.ce.berkeley.edu/~daganzo/Simulations/classes/CityApplet.html>.

idealizations that will be used in this dissertation. Then, an interactive simulation of this closed system is used to unveil evidence of chaotic network behavior when vehicles are allowed to turn between multiple routes. Chapter 3 further idealizes this closed system so that it can be described analytically. This analysis shows the emergence of multiple equilibrium states and the tendency of congestion to distribute itself unevenly across space when the network becomes congested. Chapter 4 generalizes the model presented in Chapter 3 to include entering and exiting flows in order to examine network dynamics. This analysis shows that the dynamics of a rush hour are not symmetric—network behavior is much different during the beginning of the rush hour than during the end of the rush hour. Chapter 5 presents an analytical methodology that can be used to examine how conflicting turning maneuvers between different routes on a network changes the network’s maximum efficiency. Finally, Chapter 6 summarizes the major findings of this dissertation and provides directions for future research.

Chapter 2

Simulation evidence of unstable behavior

This chapter presents simulation evidence that unveils the chaotic macroscopic behavior that arises on a simple network when vehicles turn between multiple routes. Section 2.1 describes the basic grid network that will be used for analysis in both this and the remaining chapters. Section 2.2 presents an idealization of this grid network that isolates the effects of turning maneuvers across multiple routes. Section 2.3 provides details about an interactive simulation that was created to study the idealized network. Section 2.4 describes the phenomena that arise when vehicles are allowed to turn between routes. Finally, Section 2.5 summarizes the findings of this chapter.

2.1 Basic grid network

The network considered in this dissertation is a basic grid network. For now, assume that it consists of one-way streets with alternating directionality; see Figure 2.1. All links in each family of parallel streets, indexed by i where $i = 1, 2$, have the same length, L_i , and traffic flows on each obeys the same fundamental diagram, $Q_i(k)$, where k represents the average vehicle density of the link. For simplicity, the fundamental diagrams are assumed to be triangular with capacity q_{ic} , critical density k_{ic} , free flow speed v_i and backward wave speed $w_i < v_i$ (see Figure 2.2).¹ The triangular shape is chosen because it is simple and realistic, but the qualitative behavior of the system does not hinge on this particular shape.

To maintain a fixed number of vehicles in the system, every time a vehicle leaves a street at its most downstream end, another vehicle is simultaneously inserted at the most upstream end of that street. Otherwise, vehicles are not allowed to enter or exit the links in any way other than making turning maneuvers at intersections. This set up ensures that vehicles are conserved in each individual street if vehicles are not allowed to turn.

¹The maximum density of vehicles on each link is simply $k_{ij} = k_{ic} + q_{ic}/w_i$.

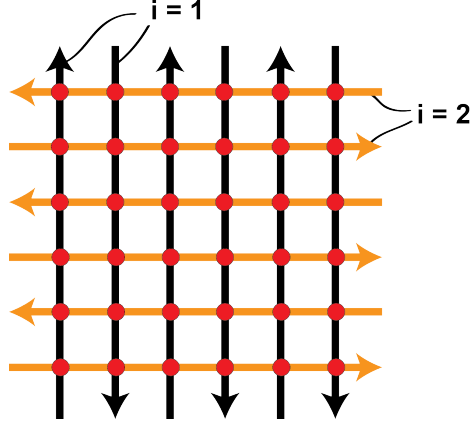


Figure 2.1. A simple grid network with two families of streets.

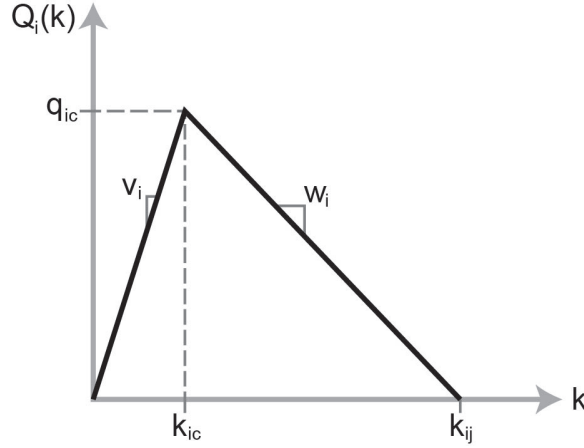


Figure 2.2. Triangular fundamental diagram of individual links.

It is assumed for now that the route taken by a vehicle is independent of traffic conditions; i.e., that drivers do not adapt in response to these conditions. Therefore, turning maneuvers can be exogenously determined. This is the same as if drivers predetermined their routes. This assumption will be relaxed later in this dissertation to examine how the presence of drivers that make routing decisions in response to congestion might change system behavior.

2.2 A two-ring idealization

To further simplify analysis, assume that the fraction of vehicles that turn from one street family to the other at any given intersection is P_T , and that this fraction is the same for all intersections at all times. This assumption, called ‘correlated turns’, ensures that if the link density profile at a moment in time is the same for all links within each family of streets, it stays the same among all links within each family

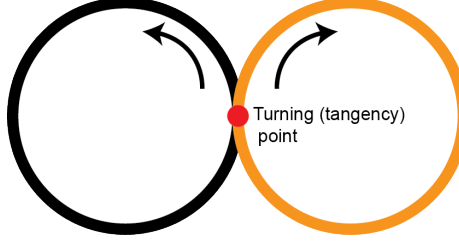


Figure 2.3. Two-ring idealization of simple grid network.

from then on. This simplification allows the system to be studied by keeping track of only two links, one from each family of streets. Therefore, a physically equivalent model of the grid network consists of two tangent rings where vehicles circulate on each (clockwise on one and counter-clockwise on the other) and are allowed to switch (or turn) between them when they reach the point of tangency (also called the turning point); see Figure 2.3. The line shadings in Figure 2.3 correspond to the two street families in Figure 2.1. This system is ideal to study because it is the simplest closed system that contains multiple overlapping routes and that allows turning between these routes.

Note that a conflict may arise in this two-ring system at the intersection when two vehicles arrive to the tangent point simultaneously and wish to travel to the same downstream location; i.e., when one vehicle wishes to change rings and the other wishes to remain on its current ring. It is assumed that when this conflict occurs, both turning and through-moving vehicles have equal priority. Therefore, one vehicle is randomly assigned priority to move and the other vehicle must yield.

2.3 Interactive simulation

An interactive simulation was created to study this idealized two-ring system. This simulation is available online and the reader may use the simulation to verify the phenomena that are presented in this chapter.²

To reduce complexity and further isolate the effects of turning between routes, it is assumed in the simulation that the two rings are identical; that is, they have the same length, $L_1 = L_2 = L$, and share the same fundamental diagram, $Q_1(k) = Q_2(k) = Q(k)$. The following parameters describe the triangular fundamental diagram used in the simulation: $v = 60$ mi/hr, $w = 15$ mi/hr, $k_c = 30$ veh/mi and $k_j = 150$ veh/mi.³ An arbitrary ring length of $L = 0.4$ mi was used in the simulation, which is equivalent to the space required by 60 vehicles at jam density.

²This simulation is available at http://www.ce.berkeley.edu/~daganzo/Simulations/two_ring_sim.html

³The subscript i is dropped from the parameter notation since the two rings are perfectly symmetric.

2.3.1 Simulation logic

The simulation was coded in the Java programming language using the CA(M) cellular automata model proposed in Daganzo (2006). As explained in this reference, the CA(M) model realistically describes traffic behavior on links because it is equivalent to the kinematic wave theory of traffic flow (Lighthill & Whitham, 1955; Richards, 1956), accurately reproducing vehicle trajectories and shock waves. It is also equivalent to the CA(L) model in Nagel & Schreckenberg (1992), but simpler to code. The model is a discrete simulation in which each ring is broken up into homogeneous cells, each the length of the jam spacing for a vehicle ($\Delta x = 1/k_j = 1/150$ mi = 35.2 ft). Vehicle locations are updated every $\Delta t = 0.4$ sec as required by the CA(M) method, where $\Delta x/\Delta t = v$.

The number of vehicles in the system at any point in time is interactively specified by the user. Only even numbers are allowed to ensure that it is always possible for each ring to contain the same number of vehicles. Vehicles are added or removed from the system at the turning point. At the start of the simulation, vehicles are loaded evenly between the two rings. If the user increases the number of vehicles while the simulation is running, new vehicles will be loaded onto whichever ring has a gap available for them to enter at the turning point. Vehicles only leave the system if the user manually decreases the number of vehicles in the system. When this is done, vehicles will exit the system as they arrive upstream of the turning point until the actual number of vehicles in the system is equal to the number specified by the user.

The simulation can also include disturbances along the ring in the form of traffic signals. To ensure a completely symmetric network, signals on both rings share the same signal settings except that signals on the left ring are offset by one-half of the signal cycle with respect to those on the right ring. If included, signals are evenly spaced along each ring, with the first signal placed at the turning point. The presence of a signal at the turning point eliminates the merging conflict.

2.3.2 Simulation interface

Figure 2.4 shows a snapshot of the simulation. The panel on the left side of the screen is used to adjust the simulation settings. The user can change the probability of turning P_T , the total number of vehicles in the system, the simulation speed, the number of signals present on each ring, and the cycle length, green time, and offset of these signals. The two leftmost buttons on the bottom panel start (or pause) and reset the simulation, respectively. The other two buttons ('L-to-R' and 'R-to-L') force a single vehicle to turn from the left ring to the right ring and from the right to left ring, respectively. These buttons work only for the next vehicle to arrive at the turning point from the relevant ring.

The vehicle locations are represented as circles on the rings at the top of the simulation interface; see the top-right of Figure 2.4. Signals are represented by colored lines that extend across the vehicles' path. The line color represents the signal phase:

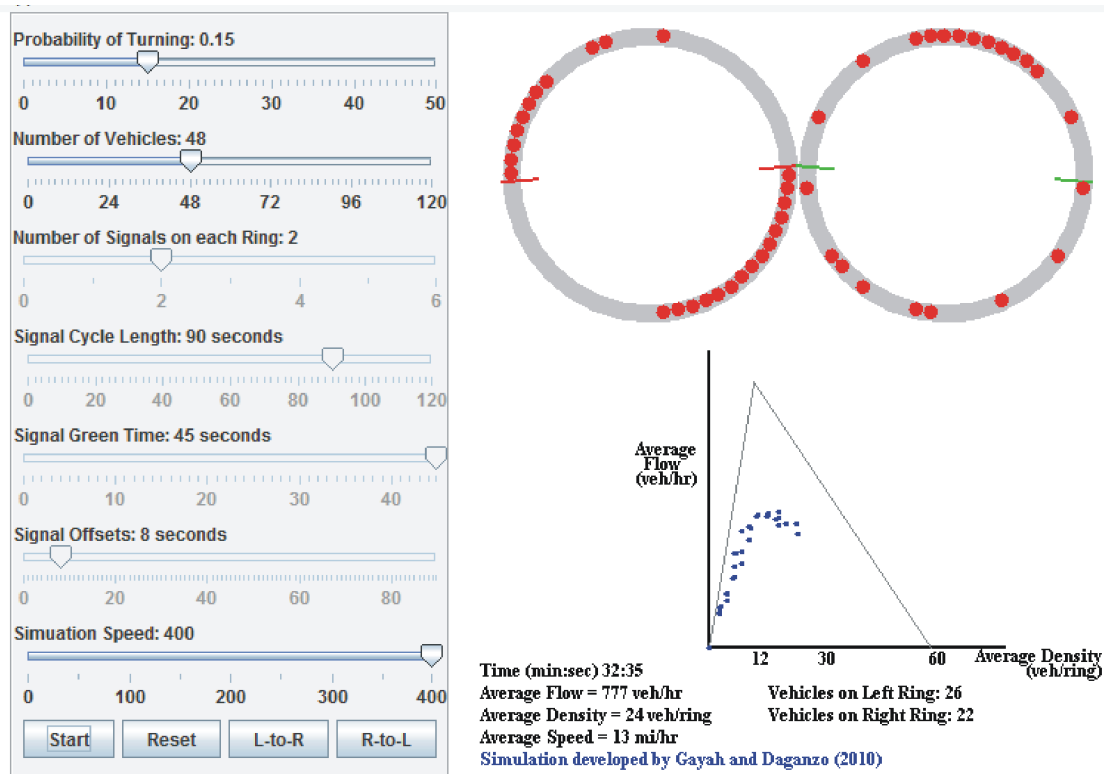


Figure 2.4. Screen shot of simulation interface showing user inputs, two-ring system and flow-density relationship.

red represents the red phase that prevents vehicle movement and green represents the green phase that allows vehicles to pass freely.

While the simulation runs, averages of flow and density taken across the entire system are calculated at discrete one-minute intervals using Edie’s generalized definitions (Edie, 1965). These observed flow-density pairs are plotted on the bottom of the simulation interface as they are calculated; see Figure 2.4. Also plotted on these axes is the fundamental diagram of the individual rings for comparison. The average flow, density and speed of vehicles in the system as well as the number of vehicles on each ring at any time are also displayed as the simulation runs.

2.4 Phenomena to observe

This section describes the macroscopic behavior of the idealized two-ring system for two cases: 1) when individual vehicles do not turn and only travel on a single ring for the entire simulation ($P_T = 0$); and, 2) when individual vehicles are allowed to turn and travel on both rings ($P_T > 0$).

2.4.1 Behavior when vehicles use a single route

When $P_T = 0$, vehicles do not change between rings as they move through the system; instead, each vehicle stays on the ring that it starts on. In this case, the two rings do not interact and can be treated independently. Since vehicles start out evenly distributed across the two rings, even distributions should remain throughout the simulation. As a result, the system should behave as predicted from theory.

For example, if no signals are present the theoretical MFD of the system predicted by Daganzo & Geroliminis (2008) is equivalent to the fundamental diagram of the rings. Therefore, observed flow-density pairs should be closely grouped along the fundamental diagram in the simulation. The two-ring simulation verifies this behavior. To confirm this, try the following steps with the online simulation: 1) set the probability of turning, number of signals on each ring, and number of vehicles to 0; 2) start the simulation; and, 3) slowly increase the number of vehicles in the system (by slowly moving the number of vehicles slider to the right by one increment every 2-3 simulation minutes) until the system becomes completely jammed. The observed flow-density pattern should be almost exactly equal to the gray fundamental diagram.

An example of such a run is presented in Figure 2.5. As expected, the observed flow-density pairs from the simulation fall exactly on the curve representing the fundamental diagram of each of the rings (i.e., the theoretical MFD of the system). This holds true even if vehicles are added to the system fairly rapidly.

2.4.2 Behavior when vehicles use multiple routes

When $P_T > 0$, vehicles are able to use both of the rings as they travel through the system, with the exact route taken varying for individual vehicles. In this case, the

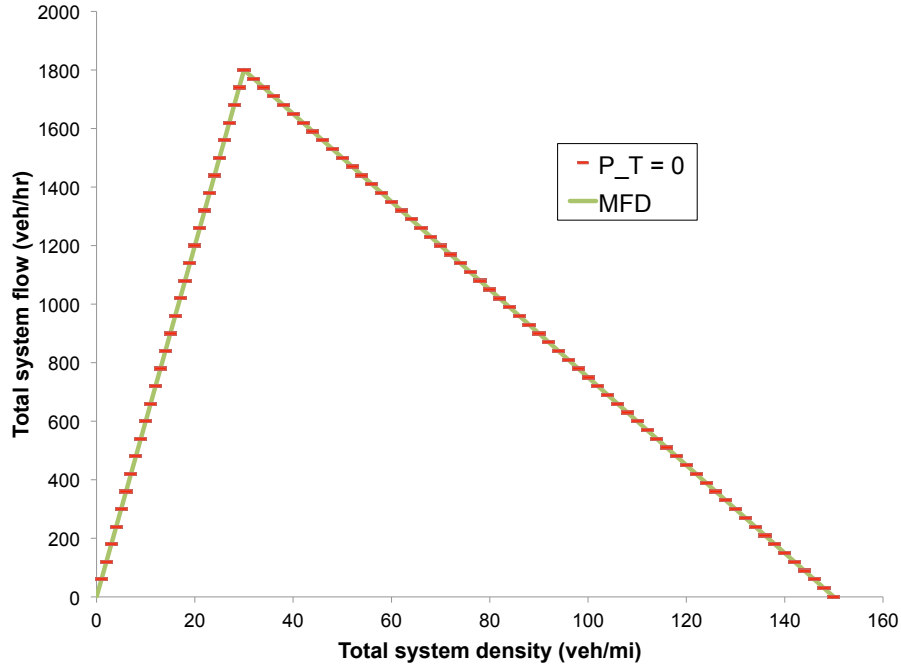


Figure 2.5. Observed flow-density relationship when vehicles do not turn.

observed flow-density pattern changes considerably from the theoretical MFD (which arises when vehicles only use a single route). To observe this change, try the following steps with the online simulation: 1) set the probability of turning to 0.05 and the number of vehicles and number of signals on each ring to 0; 2) start the simulation; and, 3) gradually increase the number of vehicles in the system until the system becomes completely jammed. Do this for different loading rates and notice that when the loading rate is very low the observed flow-density relationship closely resembles Figure 2.6.⁴

As Figure 2.6 shows, when few vehicles are in the system (e.g., total densities less than 25 veh/mi) the observed flow-density pairs are consistent with the theoretical relationship with very little scatter. The simulation shows that at these low system densities vehicles tend to distribute themselves evenly between the two rings. While some minor fluctuations in vehicle distributions may occur, these fluctuations are short-lived and vehicles gradually return to an even distribution.

However, when many vehicles are in the system (e.g., total densities greater than 30 veh/mi) the system behaves very differently. The observed flow-density pattern exhibits tremendous scatter, and observed pairs now consistently fall below the values expected from theory. For example, at a system density of 60 veh/mi the average flow varies by as much as 70% (between 300 and 1000 veh/hr) and all of the points fall below the expected value of 1400 veh/hr.

⁴The online simulation confirms that the observed flow-density pattern would be qualitatively the same for any value of $P_T > 0$.

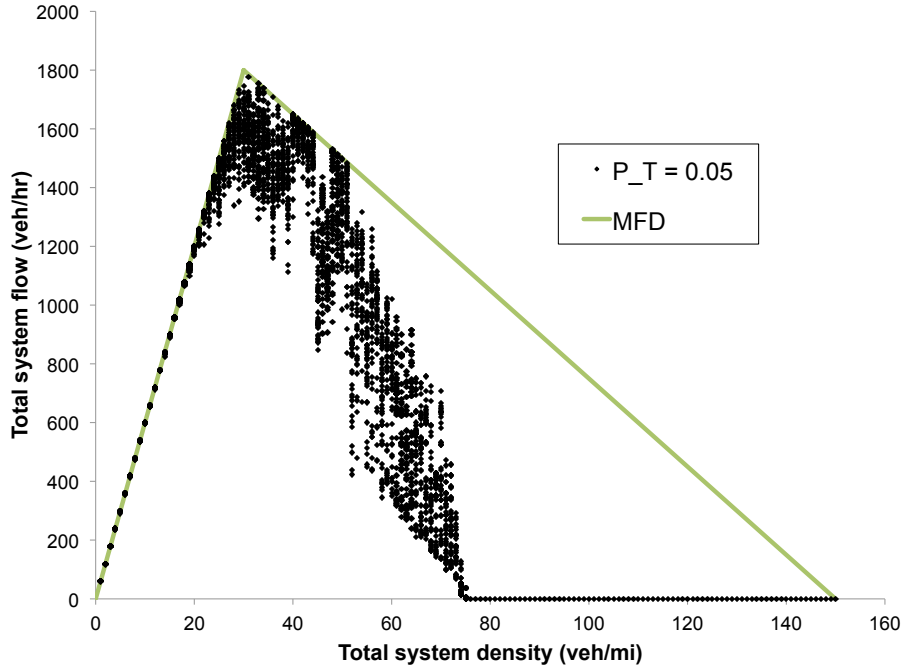


Figure 2.6. Observed flow-density relationship when vehicles do turn.

The simulation also shows that at these higher system densities vehicles now tend toward an uneven distribution—one ring will consistently have more vehicles than the other. In fact, this distribution is so uneven that the ring with more vehicles consistently operates in congestion while the other operates in free flow. Vehicles stay in this asymmetric pattern for extended periods, but the densities can occasionally flip and settle in the opposite pattern due to randomness. This imbalance occurs systematically and can be verified by repeating the simulation using the following steps: 1) set the probability of turning to 0.05, the number of signals on each ring to zero, and the number of vehicles to 40; and, 2) start the simulation. The system will start evenly loaded, with 20 vehicles on each ring. However, as the simulation runs, the average flow will change with the number of vehicles on each ring. The flow will be high (about 1,500 veh/hr) in the beginning, when each ring contains about the same number of vehicles, but will decrease as the vehicle distribution become unbalanced. For this particular example, the rings will stabilize with one containing about 33 vehicles and the other only 7. This imbalance causes the observed average flow (about 1,000 veh/hr) to be much lower than the theoretical average flow which would have been achieved if vehicles were evenly distributed between the two rings (1,500 veh/hr). In this particular example, the vehicle distribution rarely flips between rings. To observe flipping, retry this process with fewer vehicles, or a larger P_T (e.g., increase the probability of turning to 0.5), or both. When an asymmetric pattern develops, the user can try to balance the vehicle distribution manually by forcing vehicles to turn with the ‘L-to-R’ and ‘R-to-L’ buttons. However, an even distribution achieved

in this way can only be sustained temporarily. Once the user stops this process, the vehicles eventually return to an uneven distribution across the rings.

If there are enough vehicles in the system, vehicle distributions eventually become so uneven that one ring becomes completely filled with vehicles while the other remains relatively empty. When this occurs, the system completely gridlocks (or jams) and average flow reduces to zero. Once the system enters this jammed state, the imbalance remains indefinitely. Confirm this with the simulation by trying the following steps: 1) set the probability of turning to any value between 0 and 0.5, the number of signals on each ring to zero, and the number of vehicles to 60 (30 vehicles per ring, which is equivalent to 75 veh/mi or one-half of the system’s jam density); and, 2) start the simulation. Initially, the two rings will be loaded evenly (30 vehicles on each ring). As time elapses, the rings will become unbalanced, with one ring more congested than the other. Over time, the more congested ring will become more and more congested until it becomes completely jammed. Meanwhile, the other ring will be completely empty. Because of the stochastic nature of the simulation, it may take some time, but the system always reaches this jammed state. The simulation unveils that this gridlock occurs for all densities greater than 75 veh/mi, which is one-half of the theoretical jam density of the system; see Figure 2.6. When the system has an average density greater than one-half of the jam density, one ring will become completely congested, as before. However, the less-congested ring also will jam because one of the vehicles on it will try to turn onto the completely jammed ring. This vehicle will block the path of the vehicles behind it, reducing the flow on this ring to zero as well.

The behavior of the system is qualitatively the same even when traffic signals are used to resolve the conflict at the turning point, or are included along the length of the ring. An example is shown in Figure 2.7 for the case when one signal is present on each ring to resolve the conflict at the turning point, with a cycle length of 60 seconds and green time of 30 seconds. When $P_T = 0$ the observed flow-density relationship is consistent with the theory of Daganzo & Geroliminis (2008); however, when $P_T > 0$ and vehicles are allowed to use multiple routes the same phenomena are still observed: uneven vehicle distributions, scatter in the observed flow-density relationship and gridlock at lower system densities all emerge.

2.5 Major findings

The simulation of this idealized two-ring system shows that chaotic behavior is observed in traffic networks with multiple overlapping routes when vehicles have the opportunity to turn between them. When few vehicles are in the system, vehicles distribute themselves evenly across both of the routes and the observed flow-density relationship is the same as would be expected from theory. However, when many vehicles are in the system, vehicles naturally tend to distribute themselves unevenly so that one route becomes more congested than the other. These uneven distributions cause the flow-density pattern to become scattered and fall considerably below

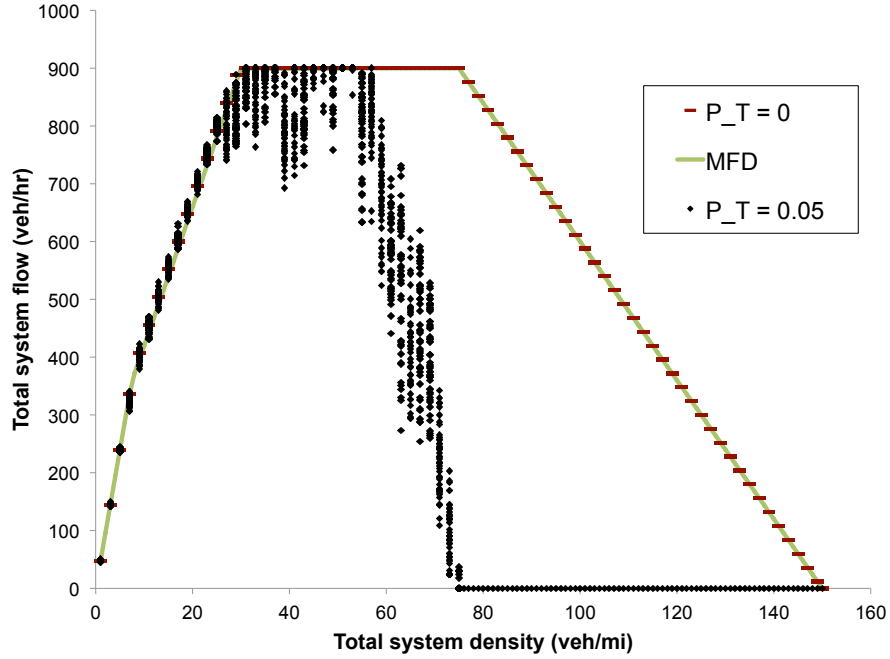


Figure 2.7. Observed flow-density relationship when signals are included, both with and without turning.

what would be expected from theory. The imbalance occurs systematically and can even cause the network to completely gridlock at relatively low system densities if there are enough vehicles to fill one route completely. Thus, multi-route systems are able to jam at densities much lower than the theoretical jam density. The simulation also confirms that this behavior occurs even if traffic signals are used to resolve the conflict at the turning point, or are included as disruptions to flow along the length of the rings.

Chapter 3

Analytical explanation of instability

Chapter 2 presented simulation evidence of the chaotic macroscopic behavior that arises in a simple traffic network when vehicles are able to turn between multiple overlapping routes. However, the simulation alone offered no insights into why turning between routes causes this behavior, or if this behavior should be expected on other types of networks. This chapter analytically studies another idealization of the grid network (which is also an idealized version of the two-ring system presented in Chapter 2) and shows that an instability arises in congested networks when vehicles are allowed to turn. Section 3.1 presents the second idealization of the grid network, which can be studied analytically, and describes the equations that govern its behavior. Section 3.2 presents an analysis of the most simple case: a homogeneous square grid network. This analysis discovers the emergence of multiple equilibrium states when the idealized system becomes congested, and confirms their existence on more realistic networks with simulation. Section 3.3 describes how the behavior of the system might change for more complicated grid networks. This includes grids that are heterogeneous, rectangular or have asymmetric demand patterns. Section 3.4 shows how the presence of drivers with adaptive routing schemes can help mitigate the instability that arises. Finally, Section 3.5 summarizes the findings of this chapter and their implications for more realistic networks.

3.1 A two-bin idealization

The two-ring model cannot be studied analytically due to the vehicle dynamics along the ring and at the turning point. To facilitate an analytical approach, further simplifying assumptions need to be made. Now assume that: (i) intersections are unsignalized and organized as in Figure 3.1; and, (ii) merging traffic (consisting of the vehicles that turn) has priority over through traffic and forces its way into the target bin unless it is jammed.

If intersections are unsignalized, a ring's density should be fairly homogeneous

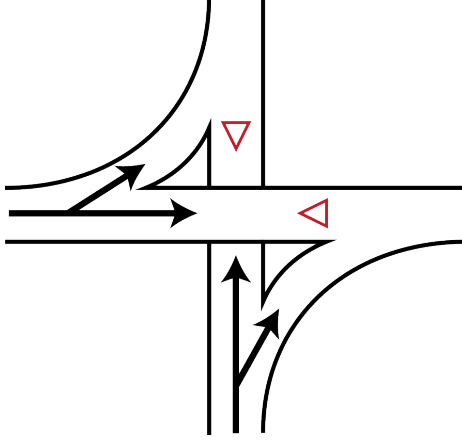


Figure 3.1. Intersection control simplification.

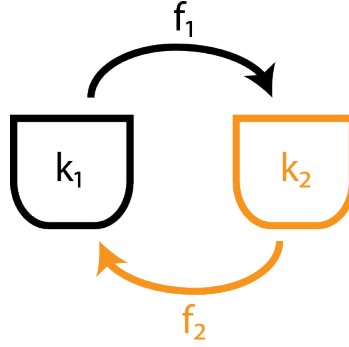


Figure 3.2. Two-bin idealization of simple grid network.

along its length near equilibrium states. Therefore, the flow at its downstream end can be approximated by:

$$q_i = Q_i(k_i) \quad (3.1)$$

where k_i is the average density of vehicles along the entire length of ring i . This simplification is convenient because it allows the state of the ring to be characterized by a single value, its average density, as if it was a memoryless ‘bin’ with an MFD Q_i instead of a ‘ring’. Thus, the state of the entire system can be described by the pair (k_1, k_2) ; see Figure 3.2. This simplification may not be realistic when conditions in the system change rapidly with time.

Also if turning traffic has priority over through traffic, the rate that vehicles leave one bin, f_i , is simply a function F_i of the average density:

$$f_i = F_i(k_i) \doteq P_T Q_i(k_i), \quad (3.2)$$

as long as neither bin is completely filled (or jammed) with vehicles (i.e., $k_i \neq k_{ij}$ where k_{ij} is the jam density of bin i).

Using Edie's generalized definitions of network flow and network density (Edie, 1965), the average density of vehicles in the system can be expressed as a function of its state:

$$k_T = \frac{k_1 L_1 + k_2 L_2}{L_1 + L_2}, \quad (3.3)$$

and the average flow of vehicles in the system can be similarly expressed as:

$$q_T = \frac{Q_1(k_1)L_1 + Q_2(k_2)L_2}{L_1 + L_2} = \frac{q_1 L_1 + q_2 L_2}{L_1 + L_2}, \quad (3.4)$$

where L_i represents the length of ring i .

3.1.1 Equilibrium and stability conditions

The dynamics of the system near equilibrium can be described by the following differential equations:

$$\frac{dk_1}{dt} = \frac{F_2(k_2) - F_1(k_1)}{L_1} \text{ and } \frac{dk_2}{dt} = \frac{F_1(k_1) - F_2(k_2)}{L_2}, \text{ if } k_i \neq k_{ij}; \text{ or} \quad (3.5a)$$

$$\frac{dk_1}{dt} = \frac{dk_2}{dt} = 0, \text{ otherwise.} \quad (3.5b)$$

Equation (3.5) shows that the density of vehicles in each bin will not change with time (and thus, the system will be in an equilibrium state) when:

$$F_1(k_1) = F_2(k_2), \text{ if } k_i \neq k_{ij} \forall i; \text{ or} \quad (3.6a)$$

$$\frac{dk_1}{dt} = \frac{dk_2}{dt} = 0, \text{ otherwise.} \quad (3.6b)$$

The stability of these equilibrium states is also of interest. Unstable equilibrium states are transient and can only be observed temporarily. However, stable equilibrium states are those that the system naturally tends towards and persist for a long time. These stable states can be used to describe the actual macroscopic behavior of the system.

In the two-bin system, equilibrium states are stable if and only if a perturbation in k_i from the equilibrium state always causes a perturbation in dk_i/dt in the opposite direction. From (3.3), a perturbation of ϵ in k_1 is accompanied by a perturbation of $-(L_1/L_2)\epsilon$ in k_2 if k_T is held constant. Approximating (3.5) with a first-order Taylor series and inserting these two perturbations reveals that:

$$\frac{dk_1}{dt} = -\frac{dk_2}{dt} = -\left[\frac{dF_1/dk_1}{L_1} + \frac{dF_2/dk_2}{L_2}\right]\epsilon. \quad (3.7)$$

Thus, an equilibrium state that satisfies (3.6a) will be stable if and only if:

$$\frac{dF_1(k_1)/dk_1}{L_1} + \frac{dF_2(k_2)/dk_2}{L_2} > 0, \text{ if } k_i \neq k_{ij} \quad (3.8)$$

If (3.8) does not hold, a small perturbation would grow with time and the system will tend away from that equilibrium state; therefore, that equilibrium state would be unstable.

These equations are now used to describe the equilibrium patterns that can be expected for the two-bin system. Section 3.2 describes the behavior of the two-bin system with identical bins (representing a homogeneous, square grid) while Section 3.3 describes the equilibrium patterns when the bins are not identical (which may represent a rectangular or a non-homogeneous grid).

3.2 Analysis of a homogeneous square grid

Considered here is the special case where the grid is perfectly square (i.e., $L_1 = L_2 = L$) and the two families of streets are identical in all relevant traffic aspects (i.e., $Q_1 = Q_2 = Q$). Since the two bins are identical, the subscript ‘i’ is dropped from the parameter notation in the remainder of this section.

In this case, the equilibrium conditions (3.6) simplify to:

$$F(k_1) = F(k_2) \Leftrightarrow Q(k_1) = Q(k_2) \text{ if } k_i \neq k_j; \text{ or} \quad (3.9a)$$

$$k_1 = k_j \text{ or } k_2 = k_j. \quad (3.9b)$$

and the stability conditions (3.8) become:

$$\frac{\partial F(k_1)}{\partial k_1} + \frac{\partial F(k_2)}{\partial k_2} > 0 \Leftrightarrow \frac{\partial Q}{\partial k_1} + \frac{\partial Q}{\partial k_2} > 0; \text{ or} \quad (3.10a)$$

$$k_1 = k_j \text{ or } k_2 = k_j. \quad (3.10b)$$

Note that (3.9) and (3.10) are independent of the turning fraction, P_T , for $P_T > 0$.

Similarly, average network density and network flow simplifies to:

$$k_T = \frac{k_1 + k_2}{2}, \quad (3.11)$$

and:

$$q_T = \frac{q_1 + q_2}{2} = \frac{Q(k_1) + Q(k_2)}{2}. \quad (3.12)$$

An analysis of the identical two-bin system can now proceed. Section 3.2.1 describes the state space of the system and how it translates onto the flow-density plane. Section 3.2.2 describes the equilibrium states of the system. And, Section 3.2.3 confirms that similar qualitative behavior is observed on a more realistic network.

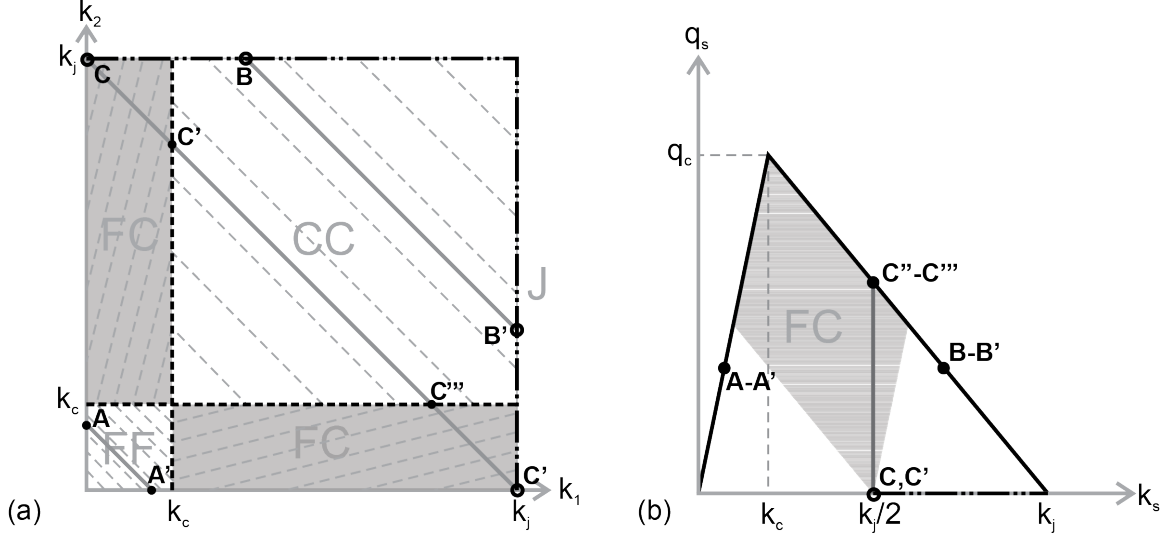


Figure 3.3. (a) Phase diagram displaying the state space of the symmetric two-bin system; and (b) plot of system flow (q_T) vs. system density (k_T) in relation to MFD of individual bins.

3.2.1 State space and flow-density relationship of homogeneous square grid

As previously mentioned, the state of the system is defined by the pair (k_1, k_2) , and this state describes traffic conditions on both bins. All possible states can be displayed graphically on a phase diagram such as the one in Figure 3.3a. At any given time the system can exist in one of four possible regimes, based on the traffic conditions in the two bins. These regimes are depicted in Figure 3.3a. The first regime (bottom-left quadrant) is the **Free-Free** or FF regime because for states in this regime both bins operate in free flow conditions. The second (top-right quadrant) is the **Congested-Congested** or CC regime because both bins operate in congestion. The third (shaded areas) is the **Free-Congested** or FC regime because one bin operates in free flow and the other in congestion. The last (denoted by the bold dotted-dashed lines) is the **Jam** or J regime because at least one of the bins is completely jammed.

Equi-flow contours are also plotted on Figure 3.3a as dashed gray lines. Flows represented by these contour lines increase towards the interior where they take the form of trapezoids. Three lines of constant density are also plotted on the phase diagram: segments $\overline{AA'}$, $\overline{BB'}$, and $\overline{CC'}$. These segments are diagonal lines with a slope of -1 for square grids. The two sets of contours can be used to define a mapping from the phase diagram to the flow-density plane: $(k_1, k_2) \rightarrow (k_T, q_T)$.¹ This mapping is shown in Figure 3.3b. Consideration shows that the two quadrants representing the FF and CC regimes are mapped onto the bold triangle that represents the theoretical MFD of the system. This is a surjective mapping in which constant-

¹This mapping can of course be expressed mathematically using (3.11) and (3.12).

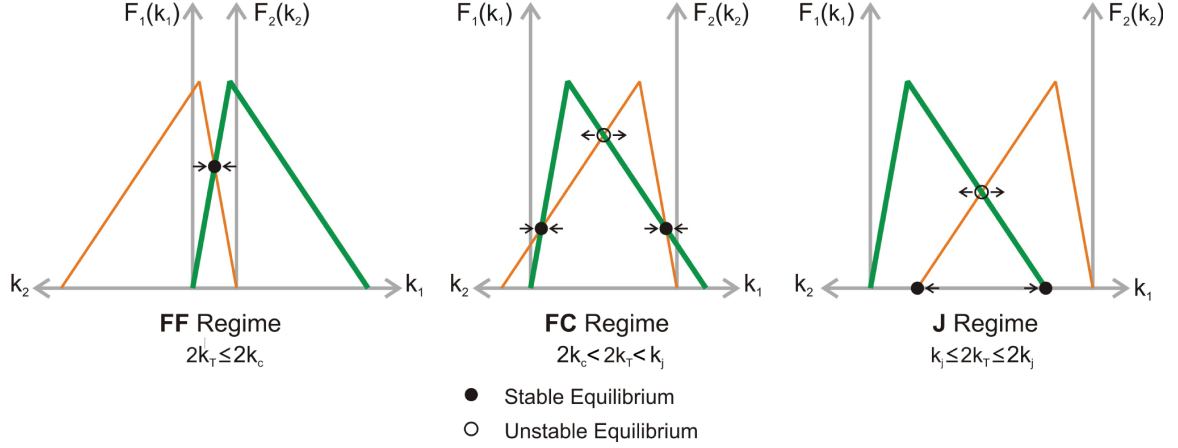


Figure 3.4. Possible regimes for symmetric two-bin system.

density segments such as $\overline{AA'}$, $\overline{BB'}$, and $\overline{C''C''''}$ are mapped to points on the flow-density plane. Therefore, in these areas average flows are dependent only on the total system density and independent on the distribution of vehicles between the two bins. The distribution of vehicles also does not matter in the J regime, which maps directly onto the bold dotted-dashed line on the flow-density plane.

In contrast, points in each one of the rectangles representing the FC regime map one-to-one with points in the shaded parallelogram on the flow-density plane. For example, line segments $\overline{CC''}$ and $\overline{C''''C''}$ are mapped onto line segments on the flow-density plane. Therefore, in these areas average flows are determined by both the total system density and the distribution of vehicles between the two bins.

3.2.2 Equilibrium behavior

Equations (3.9) and (3.10) can be used to determine the set of stable equilibrium states, and (3.11) and (3.12) can be used to calculate the expected flow-density relationship described by the set of stable equilibrium states.

Figure 3.4 shows how the set of equilibrium states can be determined for each value of k_T by superimposing plots of $F(k_1)$ (the dark curve) and $F(k_2)$ (the light curve). In each frame of this figure, the two curves are drawn with the k_1 and k_2 axes running in opposite directions. In this construction, the origins of the two curves are separated by $2k_T$ units on the x-axis and any point on the abscissa in-between the two vertical axes identifies a (k_1, k_2) pair. The (k_1, k_2) pairs that denote the intersection of the two curves satisfy (3.9a) and those that denote the jam density of either curve satisfy (3.9b); therefore, these pairs are equilibrium states.

The stability of each equilibrium state can be checked using (3.10) and it can also be verified using this graphical construction. Imagine that a small perturbation moves the density on the abscissa slightly to the left or right of any equilibrium point. If at this new state the rate at which vehicles leave bin 1 is greater than the rate at which vehicles leave bin 2 (i.e., $F(k_1) > F(k_2)$), then bin 1 experiences a net outflow—the

librium states emerge in which one bin contains more vehicles than the other; i.e., for densities greater than the bifurcation density vehicles will naturally tend toward an uneven distribution between the two bins. The set of stable equilibrium states moves from the FF regime before the bifurcation to the FC regime after. At these higher system densities, an even distribution of vehicles is no longer sustainable as any minor perturbation will cause the system to tend away from an even distribution. This continues until the set of stable equilibrium states enters the J regime where the more congested bin is completely jammed.

The flow-density relationship for the set of equilibrium states is presented in Figure 3.5b. The reproducible MFD of the system is the flow-density relationship defined by the set of stable equilibrium states since this represents the robust and reproducible relationship between average flow and average density of the system. For low system densities vehicles naturally distribute themselves between the two bins. Since stable equilibrium states for these densities are in the FF regime, the reproducible MFD of the system is equivalent to the fundamental diagram of the individual bins. However, for high system densities even distributions of vehicles are not sustainable. The set of stable equilibrium states lies in the unbalanced FC and J regimes that, as mentioned in Section 3.2.1, are associated with lower average flows than would be achieved with even vehicle distributions. The lower average flows at unbalanced stable equilibrium states (compared with balanced unstable equilibrium states) can clearly be seen in the last two frames of Figure 3.4.

The tendency toward uneven vehicle distributions in congested networks may help explain why the current macroscopic theory (which always assumes even vehicle distributions) over-predicts observed flow-density pairs in congested networks. Additionally, the existence of multiple stable equilibrium states may help explain some of the scatter observed in plots of aggregate flow vs. density. As examples, the existence of multiple equilibrium states explains why uneven vehicle distributions occasionally flip in the two-ring model, and the system periodically jumping between stable equilibrium states can explain the wide range of flows observed for a single value of density. The (stable) MFD also predicts zero flows for $k_j \leq 2k_T \leq 2k_j$ because the system tends toward the J regime and jams at densities well below k_j . The jamming of a road network at densities much lower than jam density is a phenomenon that is typically seen in simulations of networks that are more complex than the two-ring system.

Also note that these effects are independent of the proportion of turning traffic, P_T , as long as $P_T > 0$, even though they are caused by turning between different routes. Previous thinking (Daganzo & Geroliminis, 2008) assumed that small amounts of turning traffic would not have a significant effect on the distribution of traffic in a homogeneous network, and therefore have no effect on the network’s MFD. However, this simple system shows that if drivers travel using predetermined routes and do not adapt to congestion, this conjecture is false.

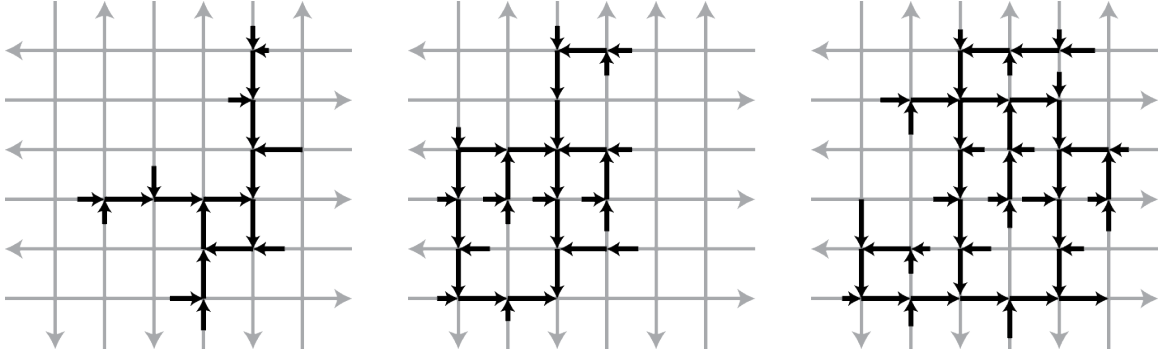


Figure 3.6. Three possible patterns of jams in a homogeneous network.

3.2.3 Tests with a more realistic network

The idealized two-bin model (and two-ring model presented in Chapter 2) isolates the effects of turning maneuvers and multiple routes to examine their effects on system behavior. These models show that, at least in some cases, turning between multiple overlapping routes can create an instability in a congested network, and that this instability may explain network-wide phenomena such as scatter in the flow-density relationship and theoretical over-prediction. However, these models are simple and the results might not be generalizable to more realistic networks.

For example, the models make the restrictive assumption that turns are perfectly correlated at all intersections in the network. With correlated turns a network can jam only if all links in one of the two families of streets jam. But if turns are not correlated, a grid network of the type studied here can jam in many different ways; see Figure 3.6 for some examples. In this figure, the black lines represent vehicle queues that can emerge when the network enters a gridlocked state. Note that the total number of vehicles in the system increases from the left-most to the right-most frame in Figure 3.6—the network can gridlock with different amounts of vehicles in the system. In fact, a homogeneous grid network will eventually jam if enough vehicles to fill four blocks are allowed to circulate in the network indefinitely. Therefore, for large networks virtually the entire range of densities would have a true stable MFD that corresponds to gridlock. However, the time required to reach these gridlocked states would be very large when there are not many vehicles circulating in the network. In practice, instead of looking for the set of stable equilibrium states, it might be more beneficial to determine the set of states that are sustainable for a reasonable analysis period, such as the length of a typical peak period.

To examine the “stable” MFD of a more realistic network, a microsimulation was used to simulate the behavior of vehicles traveling on a 6x6 grid network of two-lane streets.² In the simulation, each link was assumed to be 400 ft long and have a free flow travel speed of 30 mph. Intersections were controlled by a traffic signal that gave equal green time to the north-south and east-west roads with no offset. The number

²This simulation work was part of a collaborative effort described in Daganzo *et al.* (2011).

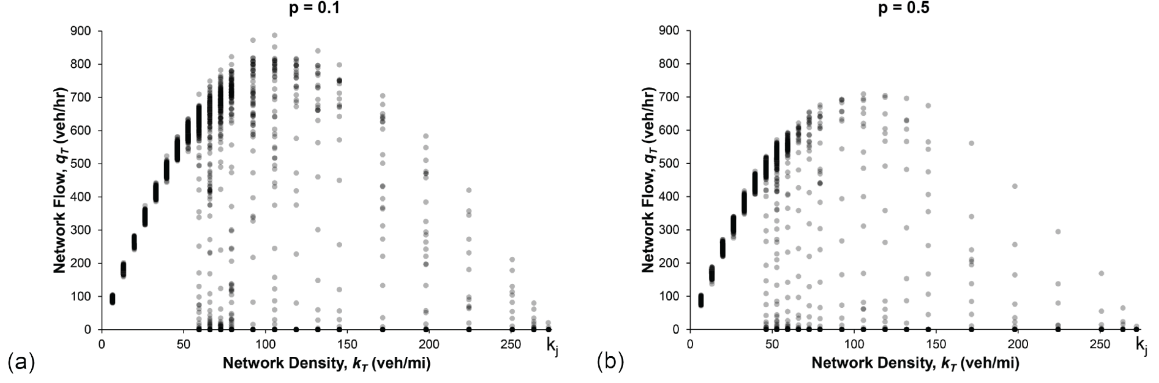


Figure 3.7. Flow-density relationship obtained from simulation of a grid network for (a) $P_T = 0.1$, and (b) $P_T = 0.5$.

of vehicles were conserved in the network by inserting a vehicle on the upstream end of a street whenever a vehicle exited at its downstream end.

The flow-density relationship obtained from putting different numbers of vehicles in the network and allowing them to circulate for 50 hours is displayed in Figure 3.7 for two different values of P_T . Each gray dot in the figure represents a one-minute aggregation of average vehicle density and average vehicle flow. Black areas indicates many overlapping observations. For low vehicle densities, the probability of gridlock is so low that gridlock is never observed, even after 50 hours. Average flows increase with density in this “stable” regime, as expected. As vehicle density increases further, however, there is a narrow metastable range where high flows are sustained for only a short period of time (on the order of a few hours). At higher densities, there is an unstable regime where the network collapses to gridlock very quickly (less than an hour). Notice that the qualitative behavior is independent of the value of P_T used. Changing P_T seems to have the biggest effect on the maximum achievable flow in the network, which will be discussed later in Chapter 5.

The main difference between the grid simulation and the two-bin model is the fact that the instability is more pronounced in the former, more realistic network. The bifurcation occurs at a much lower density, so that the increasing branch of the “stable” MFD does not persist into congestion. Thus, the grid network cannot sustain its highest flows (about 800 veh/hr for $P_T = 0.1$ and about 700 veh/hr for $P_T = 0.5$), whereas the highest flow state is sustainable in the two-bin model. This occurs because real networks can gridlock at densities less than one-half of the theoretical jam density.

3.3 Effect of network heterogeneity

This section examines the behavior of the two-bin system when the bins are not perfectly identical. This could represent networks that are not homogeneous (i.e., in which links on each of the two families of streets are governed by different fundamental

diagrams), are rectangular (i.e., in which links on each of the two families of streets do not have the same length) or have asymmetric demand patterns (i.e., in which the probability of turning is different on each family of streets).

This section may be skipped without loss of continuity. Subsequent sections in this chapter focus solely on the behavior of homogeneous square grids.

3.3.1 Non-homogeneous grid networks

In this section, the fundamental diagrams describing traffic behavior on the two families of streets are allowed to differ; see Figure 3.8 for an example in which both fundamental diagrams have the same backward wave speed and jam density.³

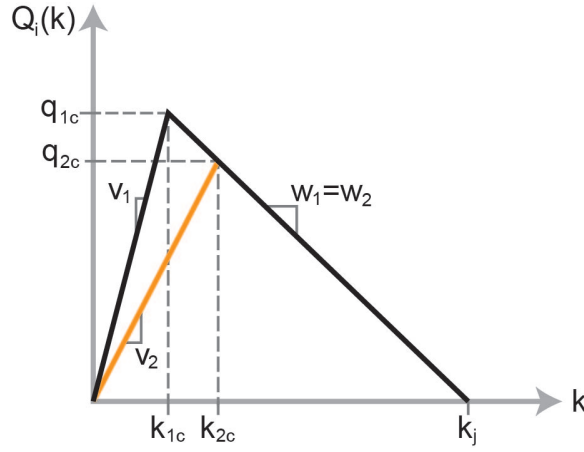


Figure 3.8. Heterogeneous triangular fundamental diagrams describing the two street families.

The equilibrium states and the stability of these states are determined using the same logic from Section 3.2, except now (3.6) and (3.8) are used instead of (3.9) and (3.10), respectively. The phase diagram and resulting flow-density relationship of all equilibrium states are presented in Figure 3.9. Notice that at low densities a single equilibrium state exists but vehicles no longer tend toward even distributions at these states. This is because vehicles will tend to distribute themselves so that the flow in each of the bins is the same, and this occurs at asymmetric vehicle distributions due to bin heterogeneity. Note also that because stable equilibrium states only exist in which the flow in each bin is the same, the maximum stable flow in the system is limited by the lowest flow in each of the bins; i.e., q_T is bounded by $\min\{q_{1c}, q_{2c}\}$. The (unstable) even distributions that are observed when both bins are congested occurs because both bins share the same backward wave speed and jam density, and is not a general result. In fact, (unstable) uneven distributions should be expected at high vehicle densities if $w_1 \neq w_2$ or $k_{1j} \neq k_{2j}$.

³This is a realistic assumption since both are properties of driver behavior.

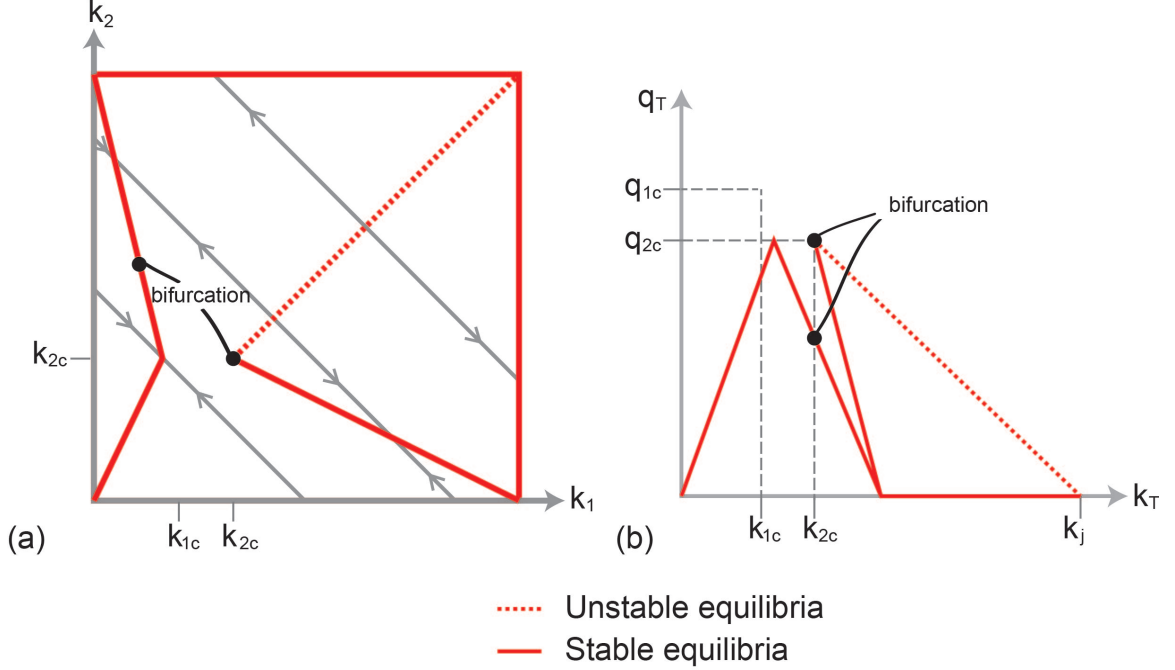


Figure 3.9. (a) Phase diagram, and (b) resulting flow-density relationship for heterogeneous two-bin system.

In addition, when the bins are asymmetric the set of equilibrium states is not continuous—two unique branches exist on the phase diagram; see Figure 3.9a. For densities lower than the bifurcation, stable equilibrium states only exist for $k_2 > k_1$; for densities higher than the bifurcation, the branch in which $k_2 < k_1$ emerges. The flow-density relationship for this heterogeneous system becomes multivalued at high densities and contains a discontinuous jump at the bifurcation due to the discontinuity in the set of stable states; see Figure 3.9b.

3.3.2 Rectangular grid networks

In this section, the fundamental diagram describing traffic for the two street families are again assumed to be identical ($Q_1 = Q_2 = Q$), and equivalent to that shown in Figure 2.2. However, the lengths of the links on each street family are now allowed to differ ($L_1 \neq L_2$); this is the two-bin model representation of a rectangular grid network. Without loss of generality, it will be assumed that $L_1 > L_2$.

The equilibrium states and their stability are again found using (3.6) and (3.8). The resulting phase diagram is identical to the phase diagram in Figure 3.5a if $L_1/L_2 < |v/w|$. However, (3.3) suggests that lines representing constant values of k_T on the phase diagram now have a slope of $-L_1/L_2$. For example, the gray lines with arrows in Figure 3.10a represent lines of constant density for the case when $L_1 = (3/2)L_2$. Because equi-density contours no longer have a slope of -1 , the mapping from the phase diagram to the flow-density plane changes. Therefore, the

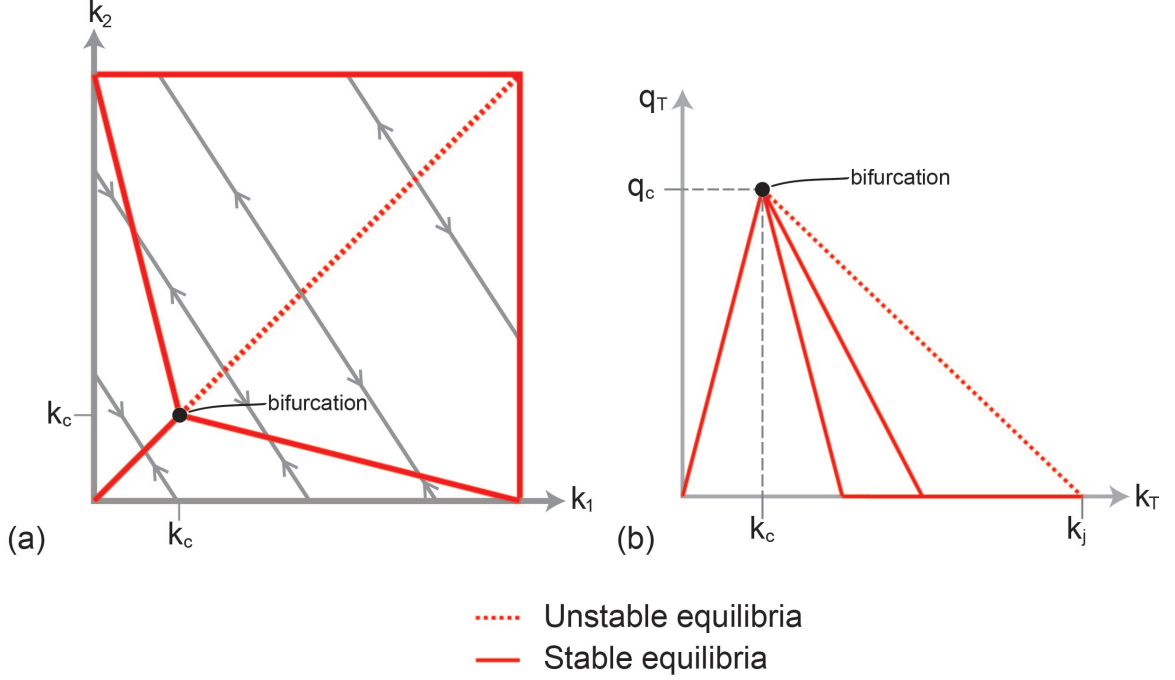


Figure 3.10. (a) Phase diagram, and (b) resulting flow-density relationship for rectangular two-bin system.

flow-density relationship for the equilibrium states will not be equivalent to Figure 3.5b, even though the set of stable states is the same on the phase diagram. Equations (3.3) and (3.4) can be used to calculate the resulting flow-density relationship for the system, as shown in Figure 3.10b. These equations show that the stable MFD is multivalued for higher vehicle densities when the network is rectangular; see Figure 3.10b.

3.3.3 Networks with asymmetric demand patterns

In this section, the bins are again assumed to be of identical length ($L_1 = L_2 = L$) and the fundamental diagrams are identical and equal to that in Figure 2.2. However, vehicles traveling on each of the two street families are now allowed to have a different probability of turning, which might represent asymmetric demand patterns within the network. The probability of an individual vehicle turning at an intersection when traveling on street family i is now denoted by P_{Ti} . Therefore, $F_i(k_i) = P_{Ti}Q_i(k_i)$.

In this case, (3.6) is satisfied when $P_{T1}Q_1(k_1) = P_{T2}Q_2(k_2)$ and (3.8) is satisfied when $P_{T1}Q_1'(k_1) + P_{T2}Q_2'(k_2) > 0$. Without loss of generality, assume that $P_{T1} > P_{T2}$. The set of equilibrium states can have two patterns based on the value of $P_{T1}/P_{T2} \doteq P_T^* \geq 1$. Figures 3.11a and 3.11b show the pattern on the phase diagram and resulting flow-density relationship for $P_T^* \leq |v/w|$ and $P_T^* > |v/w|$, respectively. Since both

$Q_1(k_1)$ and $Q_2(k_2)$ are bounded by q_c , the maximum equilibrium flow possible in the system when demand is asymmetric is:

$$q_T = \frac{q_c}{2} \left[1 + \frac{P_{T2}}{P_{T1}} \right] = \frac{q_c}{2} \left[1 + \frac{1}{P_T^*} \right]. \quad (3.13)$$

Notice that (3.13) decreases with the ratio P_T^* . As demand becomes more asymmetric, existing capacity available on the lesser used of the two street families is wasted and maximum stable flows decrease.

3.4 Effect of driver adaptation

The results so far are a ‘worst case’ because they assume that drivers follow pre-determined routes, and that they do not alter those routes due to traffic conditions. However, real drivers usually adapt their routes in response to congestion. The presence of these drivers might help smooth out the distribution of traffic in a network and eliminate or postpone the bifurcation that is observed.

To answer this question systematically, the two-bin model is now modified to study the effect of adaptive drivers. It is now assumed that some proportion, a , of drivers are ‘adaptive’, and that these adaptive drivers will modify their routes in real-time so that they do not turn from a less congested bin to a more congested one. This behavior can be modeled using the equations in Section 3.1 by simply replacing F_i with the following modified function:

$$F_i^*(k_i) = \begin{cases} (1-a)F_i(k_i) & \text{if } k_i \leq k_{i'}, \\ F_i(k_i) & \text{if } k_i > k_{i'}. \end{cases} \quad (3.14)$$

where $i' = 1, 2$ is the index for the bin being turned into.

Figure 3.12 graphically displays how F_i^* differs from F_i for three different equilibrium patterns that arise for different values of k_T when a is less than a critical value ($a_c = 1 - |w/v|$). The three different panes in Figure 3.12 each show the different equilibrium patterns that can arise in the system depending on the total system density. These patterns are again named for the character of the stable equilibrium states that exist in each. For $2k_T \leq 2k_b$ stable equilibrium states exist when the system is in the FF or CC regimes. For $2k_b < 2k_T < k_j$ stable equilibrium states exist in the FC or CC regimes. Lastly, for $k_j \leq 2k_T \leq 2k_j$ stable equilibrium states exist in the J or CC regimes.

Figure 3.13 presents the resulting phase diagram and flow-density relationship defined by equilibrium states for two cases. For levels of driver adaptation less than the critical value (Figure 3.13a), a single equilibrium state exists for k_T less than a critical bifurcation density, k_b . This equilibrium is stable and at this equilibrium state vehicles are evenly distributed between the two bins. For $k_T > k_b$, multiple equilibria emerge and the MFD defined by the set of stable equilibrium states exhibits two unique branches. This bifurcation density, k_b , increases linearly from $k_b = k_c$ when

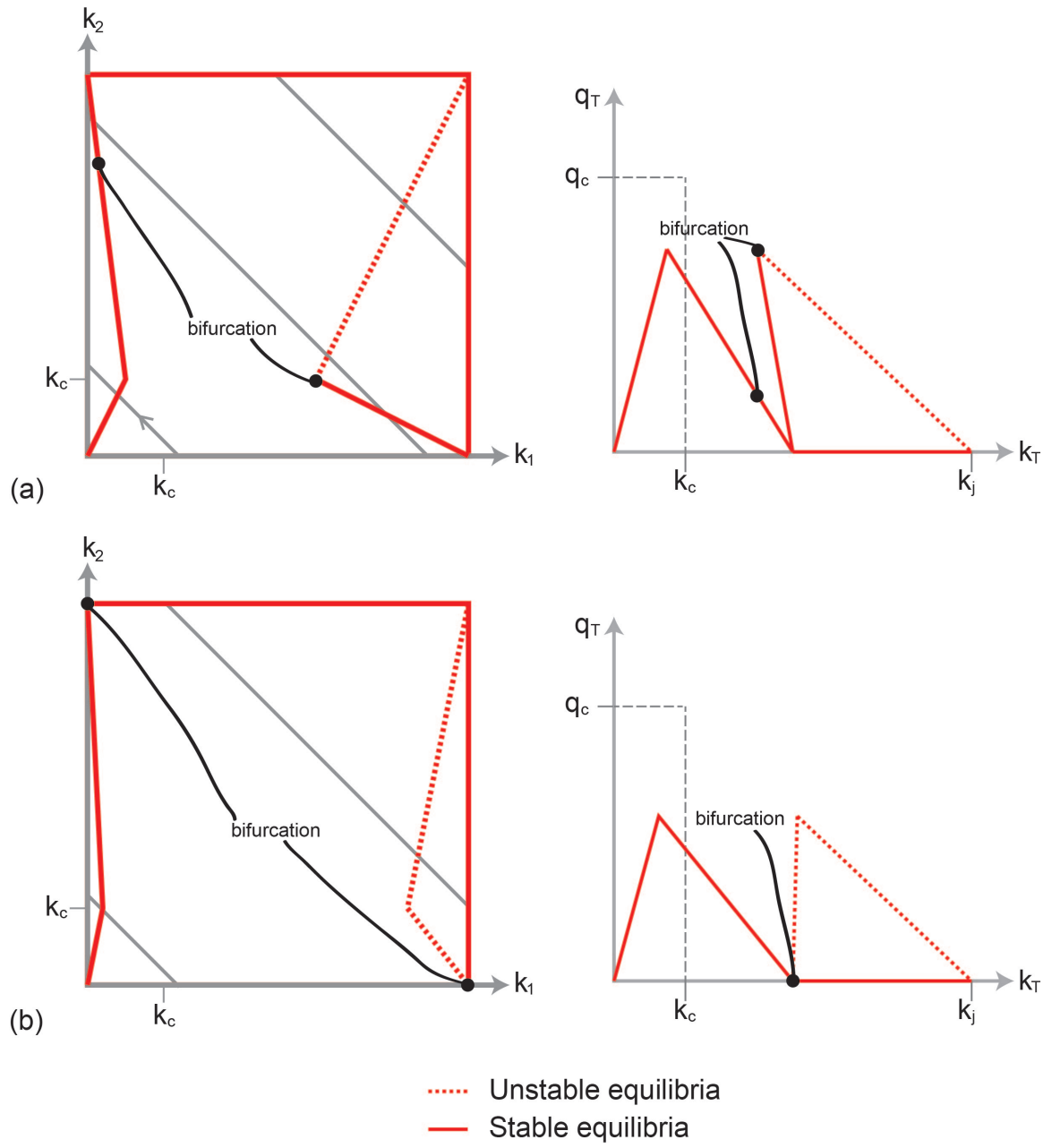


Figure 3.11. Phase diagram and flow-density relationship for asymmetric demand when: (a) $P_T^* < |v/w|$; and, (b) $P_T^* > |v/w|$.

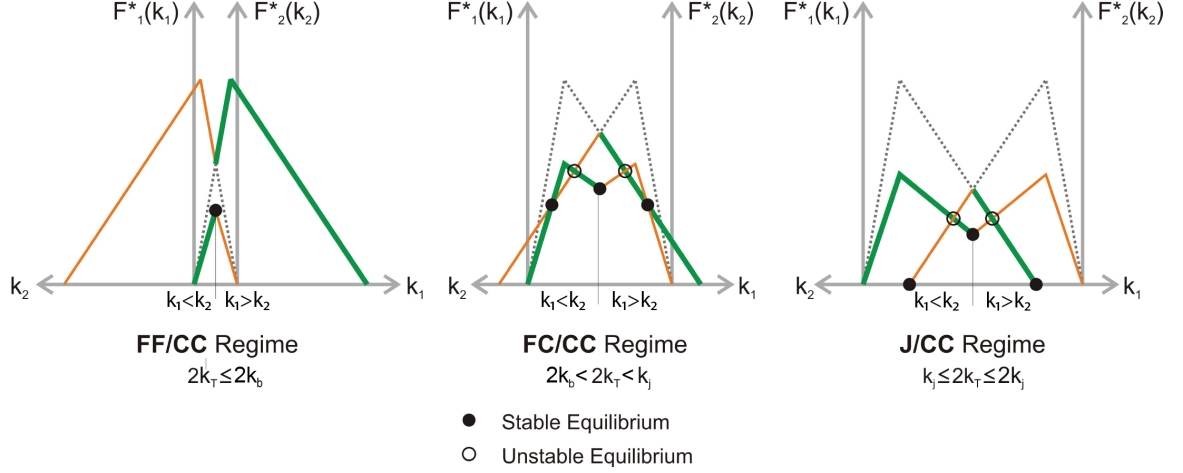


Figure 3.12. Change in $F_i(k)$ due to driver adaptation for different possible regimes when $a \leq 1 - |w/v|$. Solid lines are F_i^* and dotted lines are F_i .

$a = 0$ (no driver adaptation) to $k_b = k_j/2$ when $a = a_c$. In other words, the more adaptive drivers are, the greater the range of densities for which the MFD is single-valued. Therefore, increasing levels of driver adaptivity increases the reproducibility of the expected flow-density relationship as one might expect.

For higher levels of driver adaptation (Figure 3.13b), the lower MFD branch disappears. However, a bifurcation still takes place at $k_T = k_j/2$ and multivaluedness of the expected flow-density relationships persists for $k_T > k_j/2$.

3.5 Major findings and implications for more realistic networks

Analysis of the two-bin model shows that when few vehicles are traveling in a symmetric network, a single equilibrium state exists in which vehicles have a uniform spatial distribution. However, when the network becomes congested, a bifurcation emerges and multiple equilibrium states arise. After this bifurcation density, a uniform spatial distribution is no longer sustainable if drivers are not adaptive. In this case, the network naturally tends toward states in which vehicles are unevenly distributed. The unbalanced spatial distributions result in reduced network flows. If there is heterogeneity or asymmetric demand patterns within the network, or if the grid network is rectangular, these uneven distributions can also result in multivaluedness in macroscopic relationships between average network flow and average network density.

Fortunately, the two-bin model suggests that the presence of drivers that adaptively change routes in response to congestion helps to mitigate these uneven distributions in congested networks. When driver adaptivity is lower, the two unbalanced equilibrium states remain but even distributions also become sustainable. As driver

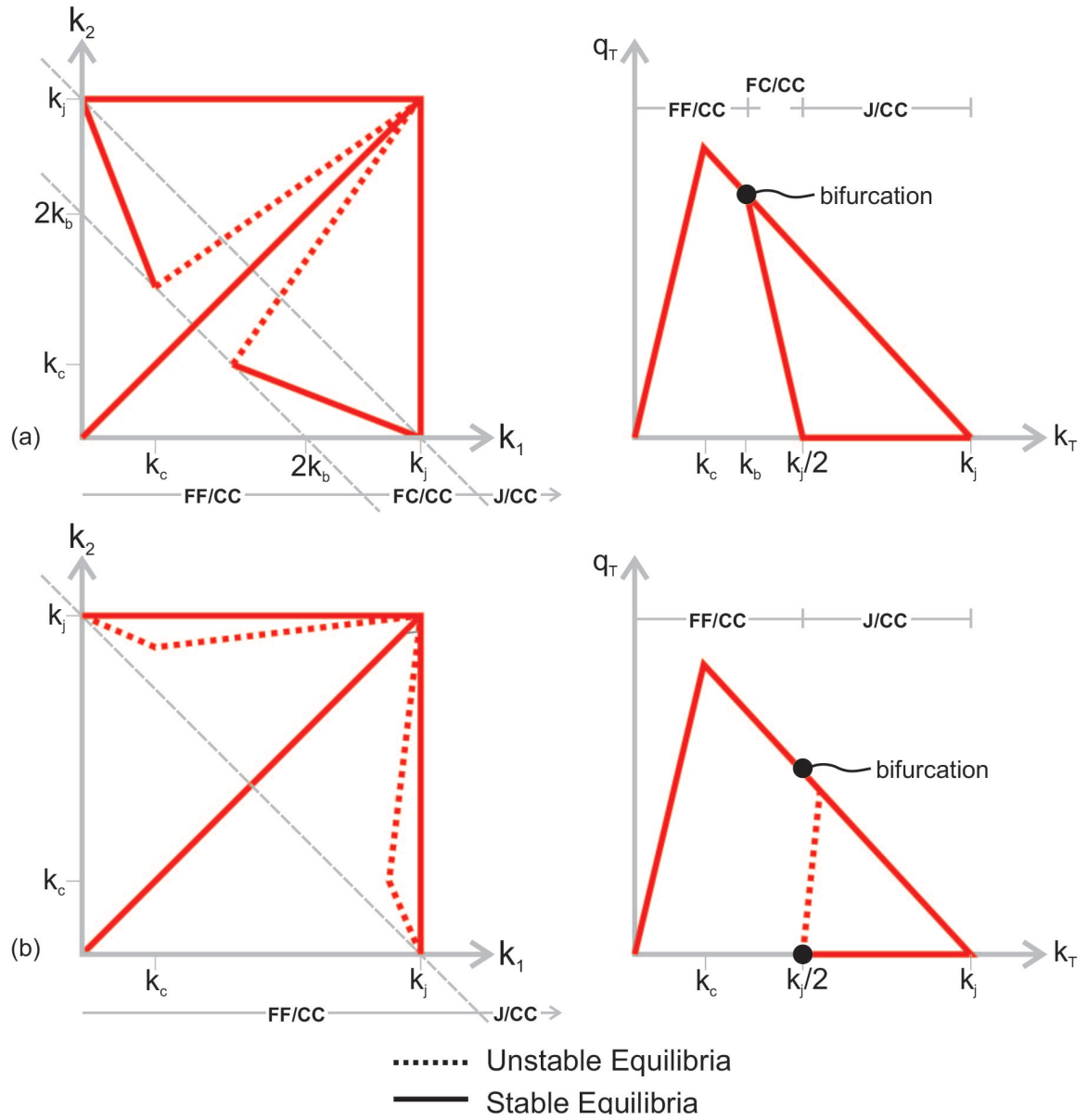


Figure 3.13. Phase diagram and MFD for two-bin model with: (a) low levels of driver adaptation ($a < a_c$); and, (b) high levels of driver adaptation ($a > a_c$).

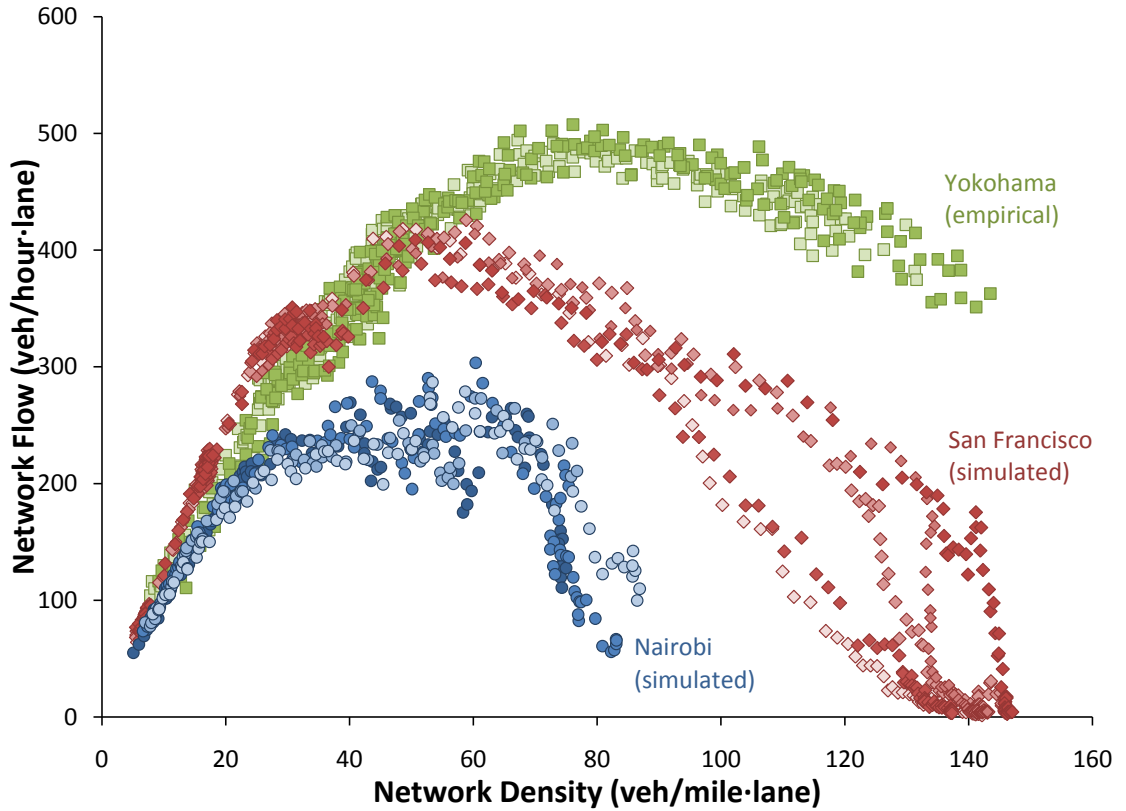


Figure 3.14. Comparison of the macroscopic fundamental diagrams of Yokohama (measured, Geroliminis & Daganzo, 2008), San Francisco (simulated, Geroliminis & Daganzo, 2007) and Nairobi (simulated, Gonzales *et al.*, 2009).

adaptivity increases, the existence of stable equilibrium states in which vehicles are unevenly distributed across the network can be eliminated.

For more realistic networks in which drivers can adaptively navigate through the network by choosing less congested routes, we would expect to see similar network-wide flow-density patterns as those produced by the two-bin model. However, the location of the bifurcation would change based on how drivers are able to navigate through the network. Figure 3.14 presents again the simulated and empirical data from Yokohama, Japan, San Francisco, California and Nairobi, Kenya which support this observation.

The Yokohama network seems to have a stable relationship between network flow and density for the entire range of densities. This is probably because real drivers adapt continuously and intelligently to traffic conditions, and are able to more actively avoid locally congested regions. Although the network gets fairly congested, as evidenced by the high observed densities, a bifurcation is not observed. It is possible

that a bifurcation does exist but only emerges at a very high density that was not observed in the field during the data collection period.

In contrast, the (simulated) San Francisco network appears to undergo a bifurcation at a density of 90 veh/mi-lane. One possible explanation for this is that the routing algorithm used in the simulation does not realistically mimic decisions made by real drivers. These simulated drivers were only able to change routing decisions at a few discrete points in time, unlike real drivers who are able to adapt continuously in response to congestion. It is conjectured that this lower adaptivity causes the bifurcation to occur at a lower density, and it emerges in the set of observable vehicle densities. The flow-density relationship for Nairobi is similar to San Francisco in that a clear bifurcation is observed at densities as low as 40 veh/mi-lane. Perhaps this network under-performs not only because simulated drivers are less able to adapt, but also because the network lacks redundancy. As explained in Gonzales *et al.* (2009), lack of network redundancy prevents drivers from shifting from congested links to nearby underutilized links because the network is not well-connected. Flows are restricted at a few key intersections that vehicles cannot avoid; therefore, drivers cannot adapt as well as in the other networks and fewer vehicles are able to jam the Nairobi network than the San Francisco and Yokohama networks.

Chapter 4

Macroscopic network dynamics during rush hour

This chapter extends the two-bin model presented in Chapter 3 to include input and output flows. This more general version of the two-bin model is used to study the dynamics of urban traffic networks during rush hour periods, and unveils that these dynamics are asymmetric. Section 4.1 explains the generalization of the two-bin model presented in Chapter 3. Section 4.2 describes network dynamics assuming drivers are not adaptive, and explores the asymmetry between the beginning and end of a rush hour period. Section 4.3 examines the patterns that may emerge on macroscopic traffic plots during a full rush hour cycle. Section 4.4 examines how network dynamics might change if drivers choose routes adaptively to avoid congestion. Finally, Section 4.5 summarizes the major findings of this chapter and their implications for more realistic networks.

4.1 Generalization of two-bin system

To examine the dynamics of urban traffic networks, a more general version of the two-bin model originally described in Chapter 3 is used. While each of the bins were previously assumed to represent one of the two parallel street families, they can also represent adjacent neighborhoods, as illustrated in Figure 4.1. Small neighborhoods like this can be represented by a bin of vehicles with average flows governed by an MFD, $Q_i(k_i)$, if traffic within each of the neighborhoods is assumed to be always evenly distributed in space. In this representation, L_i will refer to the total length of streets within neighborhood i .

In this chapter, the two bins will be assumed to be identical (i.e., $Q_1 = Q_2 = Q$ and $L_1 = L_2 = L$) to eliminate potential confounding factors. For consistency, $Q(k_i)$ is assumed to be of the form previously given in Figure 2.2.

Each bin i is assumed to experience an inflow, e_i , and outflow, x_i , of vehicles; see Figure 4.2. Inflows are assumed to either be constant or a function of the density of

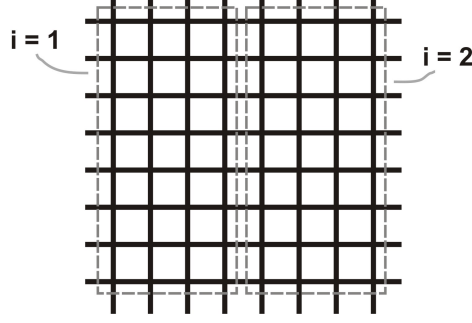


Figure 4.1. Neighborhoods within a city network that can be represented by bins.

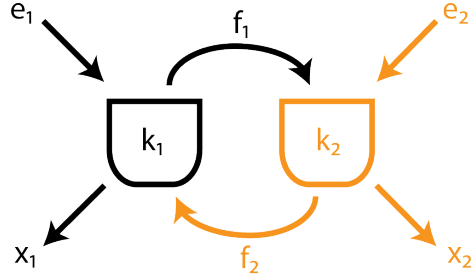


Figure 4.2. Two-bin idealization with inflows, e_i , and outflows, x_i .

vehicles in the bin entered, $e_i = E(k_i)$. Outflows will assumed to be a function of the density in the bin exited, $x_i = X(k_i)$.

4.2 System dynamics

The dynamic equations describing the behavior of the two-bin system are:

$$\frac{dk_1}{dt} = \frac{e_1 - x_1 + P_T[Q(k_2) - Q(k_1)]}{L} \text{ and} \quad (4.1a)$$

$$\frac{dk_2}{dt} = \frac{e_2 - x_2 + P_T[Q(k_1) - Q(k_2)]}{L}, \text{ if } k_1, k_2 < k_j; \text{ and,} \quad (4.1b)$$

$$\frac{dk_1}{dt} = \frac{dk_2}{dt} = 0, \text{ if } k_1 \text{ or } k_2 = k_j. \quad (4.1b)$$

These equations are now examined to determine how the system evolves during an idealized rush hour period. Section 4.2.1 describes system evolution when input flows are positive and exit flows are zero, which mimics the beginning of a rush hour period. Section 4.2.2 describes system evolution when input flows are zero and exit flows are positive, which mimics the end of the rush hour period. The more general case in which both input and output flows are positive is also described in Section 4.2.3.

4.2.1 The loading process

In this section, $e_i > 0$ and $x_i = 0$. The next subsection describes the evolution of the system in the special case when loading rates are exogenous (i.e., independent of local traffic conditions), as if entering vehicles forced their way into the traffic stream from driveways or access streets. The following subsection discusses how system behavior might change if loading is endogenous (i.e., a function of local traffic conditions), recognizing that fewer opportunities are available for vehicles to enter a traffic stream as it becomes more congested.

Balanced, exogenous loading

The behavior of the system is now examined when it is loaded uniformly at a constant rate, independent of traffic conditions within each of the bins ($e_1 = e_2 = E$). To examine the dynamics of the system, it will be convenient to denote the densities of the less congested and more congested bin by k and K , respectively; i.e., $k = \min(k_1, k_2)$ and $K = \max(k_1, k_2)$ and $k \leq K$. In this case, (4.1a) simplifies to:

$$\frac{dk}{dt} = \frac{E + P_T[Q(K) - Q(k)]}{L} \text{ and } \frac{dK}{dt} = \frac{E + P_T[Q(k) - Q(K)]}{L}, \text{ if } k, K < k_j. \quad (4.2)$$

It is interesting to determine whether the system converges to or diverges from a balanced state (i.e., a state in which $k = K$) as the average density in the system changes. From (4.2), the change in density distribution with time is:

$$\frac{d(k - K)}{dt} = \frac{2P_T[Q(K) - Q(k)]}{L}, \quad (4.3)$$

while the change in average system density with time is:

$$\frac{dk_T}{dt} = \frac{E}{L}. \quad (4.4)$$

Equation (4.4) confirms that the average system density increases with time at a fixed rate as vehicles enter the system. Since this is the case, a time-independent measure $\tau_L \doteq d(k - K)/dk_T$ can be used to describe the rate at which the density distribution changes with (increasing) system density. Note that τ_L is simply the ratio of (4.3) and (4.4):

$$\tau_L \doteq \frac{d(k - K)}{dk_T} = \frac{2P_T}{E} [Q(K) - Q(k)]. \quad (4.5)$$

The system converges toward a balanced state if $\tau_L > 0$ (i.e., if $d(k - K)$ is positive) and diverges if $\tau_L < 0$ (i.e., if $d(k - K)$ is negative). Equation (4.5) shows that if loading is exogenous and evenly distributed between the two bins the system will converge towards a balanced state if $Q(K) > Q(k)$ and diverge from a balanced state if $Q(K) < Q(k)$. These areas of convergence and divergence are displayed in

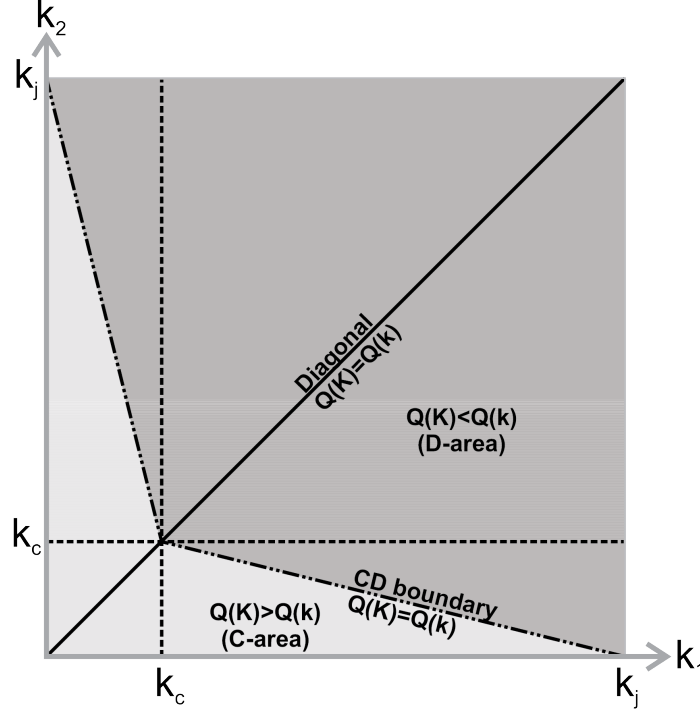


Figure 4.3. Areas of convergence to and divergence from even vehicle distributions for balanced, exogenous loading and recovery.

Figure 4.3 as the C-area and D-area, respectively. Note from (4.5) that the magnitude of the convergence measure depends on the ratio P_T/E ; the lower the rate that vehicles enter the system and the greater the probability of turning, the greater the tendency of the system to converge or diverge.

Figure 4.4a presents the evolution of the system on the phase diagram according to (4.2) for various starting points; these lines are known as ‘loading paths’. As indicated by (4.5), the loading paths converge upon the diagonal in the C-area and diverge away from the diagonal in the D-area. Also, as expected from (4.2) and (4.5), if the system starts at a balanced state it will remain balanced throughout the loading process. Moreover, if the system starts at an unbalanced state then it remains unbalanced and will never reach the diagonal. Note also that the individual loading paths are closest to the diagonal when they cross the CD boundary, which is the set of points for which $Q(K) = Q(k)$.

Also note that the converging and diverging behavior in the system for the case of balanced, exogenous loading is caused by the turning that occurs between the two bins. If turning is non-existent (i.e., if $P_T = 0$) then $\tau = 0$ and any imbalance between the two bins will not change with time. Loading paths for this special case are presented in Figure 4.4b.

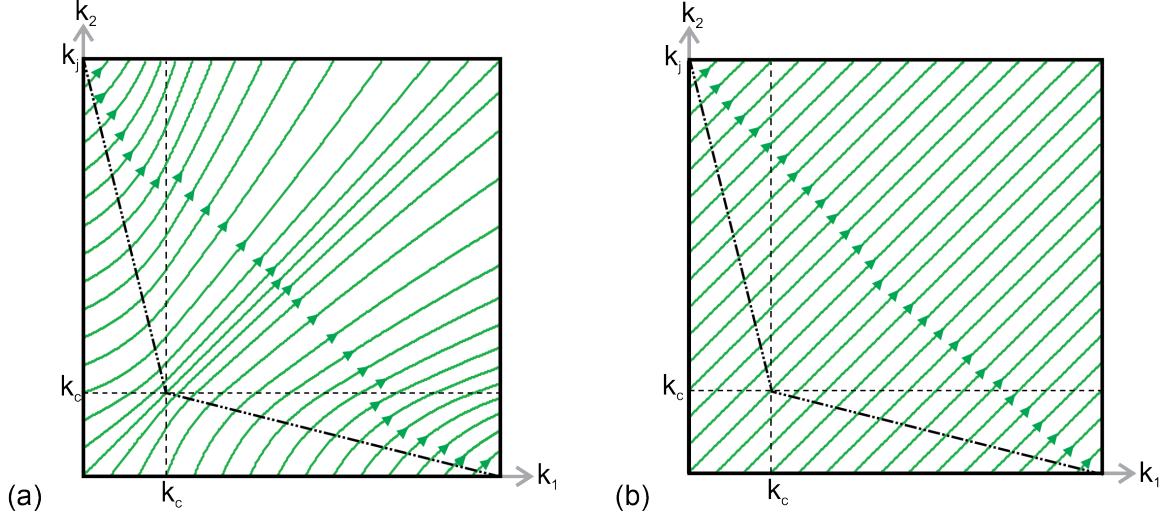


Figure 4.4. Loading paths for exogenous loading when $E = 0.2q_c$ and: (a) $P_T = 0.05$; and, (b) $P_T = 0.0$.

Endogenous loading

This section examines how the loading process might change if the rate at which vehicles enter the network is endogenously determined as a function of the density of the bin that is entered. This section can be skipped without loss of continuity; subsequent sections in this chapter focus solely on the case in which entering flows are exogenous and balanced.

If vehicles are not allowed to force their way into the traffic stream, then they must wait for an appropriate gap between vehicles before entering the network. Since these gaps become less frequent as the network becomes more congested, it is reasonable to assume that fewer vehicles may be able to enter a more congested bin. To examine how this might change the qualitative behavior of the system, assume now that $e_i = E(k_i)$. Note that because vehicles enter more congested bins at a lower rate, $E(k_i)$ is a monotonically decreasing function with respect to k_i .

The dynamic equations now become:

$$\frac{dk}{dt} = \frac{E(k) + P_T[Q(K) - Q(k)]}{L} \text{ and } \frac{dK}{dt} = \frac{E(K) + P_T[Q(k) - Q(K)]}{L}, \text{ if } k, K < k_j. \quad (4.6)$$

From these equations, the change in density distribution with time is:

$$\frac{d(k - K)}{dt} = \frac{E(k) - E(K)}{L} + \frac{2P_T[Q(K) - Q(k)]}{L}, \quad (4.7)$$

and the change in average system density with time is:

$$\frac{dk_T}{dt} = \frac{E(k) + E(K)}{2L}. \quad (4.8)$$

Like the previous case, the ratio of (4.7) and (4.8) describes how the density distribution changes as vehicles are added to the system:

$$\tau'_L \doteq \frac{d(k - K)}{dk_T} = \frac{2[E(k) - E(K)]}{E(k) + E(K)} + \frac{4P_T}{E(k) + E(K)} [Q(K) - Q(k)]. \quad (4.9)$$

Equation (4.9) shows that τ'_L will be positive (and the system will converge to a balanced state) if $[E(k) - E(K)] + 2P_T[Q(K) - Q(k)] > 0$. Therefore, the system will converge for states in which $2P_T[Q(K) - Q(k)] > E(K) - E(k)$. The right-hand side of this condition is negative since $E(k_i)$ is a monotonically decreasing function and $k \leq K$ by definition. Therefore, the set of states for which the system will converge towards the diagonal will be larger than the C-area shown in Figure 4.3 for the case when loading rates are exogenous. Additionally, for states in which $Q(K) > Q(k)$, the system will have an even stronger tendency toward an even distribution of vehicles than before. Therefore, the existence of endogenous loading does not change the qualitative behavior of the system; however, it causes the system to tend more strongly towards even distribution of vehicles and increases the range of states for which convergence occurs.

Note that when loading rates are endogenously determined the system will always tend towards balanced states if vehicles are not allowed to turn. Equation (4.5) shows that if $P_T = 0$ then $\tau_L \geq 0$ for all states, and traffic unevenness always decreases with time. The physical reason for this behavior is that fewer vehicles are able to enter a more congested bin, so less congested bins fill up more quickly than more congested bins. This naturally reduces the congestion imbalance in the system over time. The instability introduced by turning traffic diminishes this natural tendency toward convergence when the network is congested.

4.2.2 The recovery process

Considered in this section is the evolution of the system when vehicles are allowed to exit ($x_i > 0$) without any input flows ($e_i = 0$). Theoretical (Daganzo, 2007) and empirical evidence (Geroliminis & Daganzo, 2008) suggest that the rate at which vehicles are allowed to leave a neighborhood is a fixed proportion of the flow of vehicles in that neighborhood. Therefore, exit flows are endogenously determined by a function $x_i = X(k_i) = P_X Q(k_i)$, where P_X is the proportion of the circulating flow that exits.

During recovery, (4.1a) becomes:

$$\frac{dk}{dt} = \frac{P_T Q(K) - (P_T + P_X) Q(k)}{L} \text{ and } \frac{dK}{dt} = \frac{P_T Q(k) - (P_T + P_X) Q(K)}{L}. \quad (4.10)$$

From (4.10), the change in density distribution with time is:

$$\frac{d(k - K)}{dt} = \frac{2P_T + P_X}{L} [Q(K) - Q(k)], \quad (4.11)$$

and the change in average system density with time is:

$$\frac{dk_T}{dt} = -\frac{P_X q_T}{L}. \quad (4.12)$$

These expressions are very similar to (4.3) and (4.4) except that now dk_T/dt is negative. As expected, the average density of the system now decreases with time, and decreases at a rate that is a constant proportion of the average circulating flow. In view of this, the time-independent measure of the system's tendency to converge during recovery is defined as $\tau_R = -d(k - K)/dk_T$. This can be calculated by dividing (4.11) by the negative of (4.12) and reduces to:

$$\tau_R \doteq -\frac{d(k - K)}{dk_T} = \frac{2P_T + P_X}{P_X q_T} [Q(K) - Q(k)]. \quad (4.13)$$

Note the only difference between (4.5) and (4.13) is the coefficient of $[Q(K) - Q(k)]$. Thus during recovery the system has the same C- and D-areas as in Figure 4.3. The coefficient $(2P_T + P_X)/(P_X q_T)$ shows that the greater the turning fraction and the smaller the exiting fraction, the greater the tendency to converge or diverge.

The shape of recovery paths is different than the shape of the loading paths because the coefficient of $[Q(K) - Q(k)]$ is state dependent; Figure 4.5a shows an example for $P_X = 0.2$ and $P_T = 0.05$. Of course, the direction of these paths are opposite of those in Figure 4.4a because the sign of (4.11) is different from (4.3). As in the loading case, if the system starts recovering from a balanced state, it remains balanced throughout recovery, and if it starts unbalanced it remains unbalanced. Unlike the loading case, in which the loading paths cross the CD boundary at their closest point to the diagonal, the recovery paths cross the CD boundary when they are farthest from the diagonal. Thus, if the system begins recovery from an unbalanced state in the D-area, the unevenness grows until the phase path crosses the CD boundary and enters the C-area, after which the unevenness started to diminish.

Another key difference between loading and recovery can be observed by examining system behavior when $P_T = 0$. In this case, (4.13) shows that $\tau_R \neq 0$ if $Q(K) \neq Q(k)$ so the system still has tendencies to converge to even distributions when in the C-area, or diverge from even distributions when in the D-area. The physical reason for this phenomenon is that with endogenously determined exit flows more congested areas clear more slowly than less congested areas, and this can exacerbate congestion imbalance in the system with time when the system is in the D-area. However, when loading is exogenous, (4.5) shows that $\tau_L = 0$ and there is no converging or diverging behavior; congestion imbalances do not change with time. When loading is endogenous, (4.9) shows that $\tau'_L > 0$ so that the system only has the tendency to converge toward even distributions; vehicles enter more congested areas less quickly and this reduces congestion imbalances. This is why recovery is inherently a more unstable process than loading (both when loading rates are exogenously or endogenously determined). The presence of turning creates an instability that can exacerbate the difference.

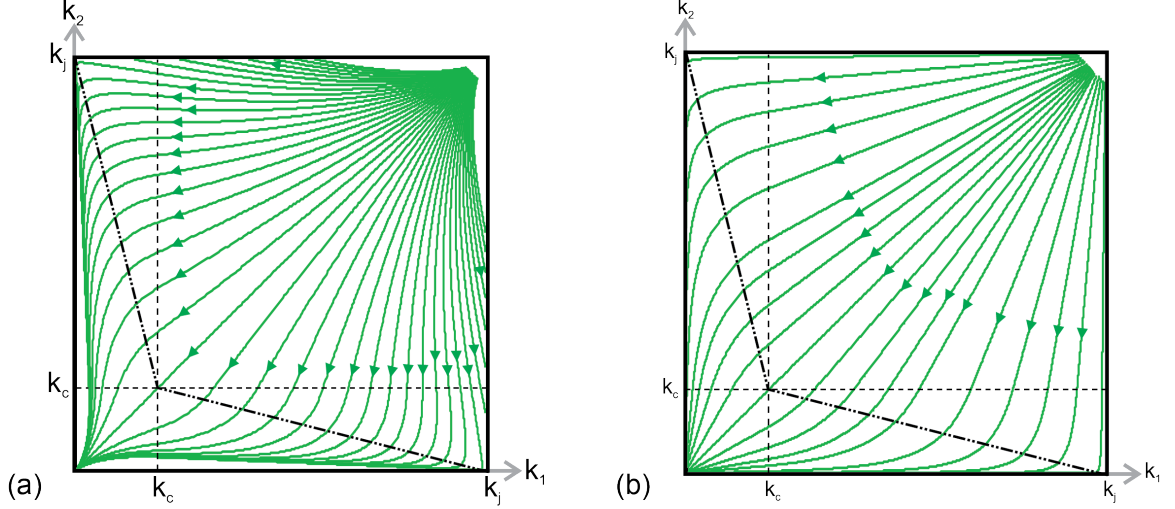


Figure 4.5. Recovery paths for $P_X = 0.2$ and: (a) $P_T = 0.05$; and, (b) $P_T = 0.0$.

4.2.3 Simultaneous entering and exiting

Here the behavior of the system is examined when vehicles are allowed to enter and exit simultaneously. The objective is to see if equilibrium states exist in which the density of the system does not change with time. For simplicity, it is assumed that vehicle entry rates are exogenous (i.e., $E_1 = E_2 = E$) and vehicle exit rates are endogenously determined as a fixed proportion of the flow in each bin (i.e., $X(k_i) = P_X Q(k_i)$).

In this case, (4.1) becomes:

$$\begin{aligned} \frac{dk_1}{dt} &= \frac{E + P_T Q(k_2) - (P_T + P_X) Q(k_1)}{L} \text{ and} \\ \frac{dk_2}{dt} &= \frac{E + P_T Q(k_1) - (P_T + P_X) Q(k_2)}{L}, \text{ if } k_1, k_2 < k_j. \end{aligned} \quad (4.14a)$$

The sum of the two right-hand sides in (4.14a) is $(2E - P_X[Q(k_1) + Q(k_2)]) / L$, while their difference is $(2P_T + P_X)[Q(k_2) - Q(k_1)] / L$. At equilibrium, both of these quantities have to be zero. Thus, an equilibrium with $k_1, k_2 < k_j$ will only occur if the second quantity is zero, i.e.:

$$Q(k_1) = Q(k_2) \doteq Q; \quad (4.15a)$$

and if in addition the first quantity is zero, i.e.:

$$Q = E / P_X. \quad (4.15b)$$

Since $Q \leq q_c$, equilibrium solutions only exist for $E / P_X \leq q_c$.

Consideration shows that there are four (k_1, k_2) that satisfy (4.15) for each value of $Q < q_c$: one in the FF regime, one in the CC regime, and one in each of the rectangles of Figure 3.3a representing the FC regime. The locus of all these equilibria for all $Q \leq q_c$ turns out to be the union of the diagonal and CD boundary of Figure 4.3. These are also the same as the union of stable and unstable equilibria on the phase diagram of Figure 3.5a.

These equilibrium solutions (k_1^{eq}, k_2^{eq}) are stable if and only if small perturbations to the equilibrium densities ϵ_1 and ϵ_2 shrink with time. Again, stable equilibria are important because the system tends towards them and they are reproducible. The reader can verify that (4.14a) can be written in terms of the perturbations as follows:

$$\frac{d\epsilon_1}{dt} = \frac{P_T Q'(k_2^{eq})\epsilon_2 - P_T Q'(k_1^{eq})\epsilon_1 - P_X Q'(k_1^{eq})\epsilon_1}{L}, \quad (4.16a)$$

$$\frac{d\epsilon_2}{dt} = \frac{P_T Q'(k_1^{eq})\epsilon_1 - P_T Q'(k_2^{eq})\epsilon_2 - P_X Q'(k_2^{eq})\epsilon_2}{L}, \quad (4.16b)$$

which can be rewritten in matrix form as:

$$\begin{bmatrix} d\epsilon_1/dt \\ d\epsilon_2/dt \end{bmatrix} = \frac{1}{L} \begin{bmatrix} -(P_T + P_X)Q'(k_1^{eq}) & pTQ'(k_2^{eq}) \\ P_T Q'(k_1^{eq}) & -(P_T + P_X)Q'(k_2^{eq}) \end{bmatrix} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} = \underline{\underline{M}} \underline{\epsilon}. \quad (4.17)$$

Clearly, the perturbations $\underline{\epsilon}^{(1)}$ after a single time step, dt , are related to the initial perturbation $\underline{\epsilon}^{(0)}$ by:

$$\underline{\epsilon}^{(1)} = [\underline{\underline{I}} + \underline{\underline{M}}dt] \underline{\epsilon}^{(0)}. \quad (4.18)$$

And, after n time steps the perturbations $\underline{\epsilon}^{(n)}$ are:

$$\underline{\epsilon}^{(n)} = [\underline{\underline{I}} + \underline{\underline{M}}dt]^n \underline{\epsilon}^{(0)}. \quad (4.19)$$

Thus, the equilibrium is stable if and only if $[\underline{\underline{I}} + \underline{\underline{M}}dt]^n \rightarrow \underline{0}$; i.e., if all the eigenvalues of $[\underline{\underline{I}} + \underline{\underline{M}}dt]$ are less than one in absolute value. Examination of (4.17) shows that all equilibrium states on the CD boundary (in the FC regime) and on the congested part of the diagonal (in the CC regime) are unstable. Only a portion of the diagonal in the FF regime is stable. This is different from the result in Chapter 3, which finds that solutions on the CD boundary are stable when a fixed number of vehicles are in the system. However, the same result as in Chapter 3 would be obtained if exiting flows were exogenous and equal to the rate at which vehicles enter the system. This further illustrates the destabilizing effect of endogenous exiting flows.

4.3 Patterns in the flow-density relationship

It is clear from Sections 4.2.1 and 4.2.2 that loading and recovery paths are asymmetric. The system tends more strongly towards evenness as vehicles enter the system

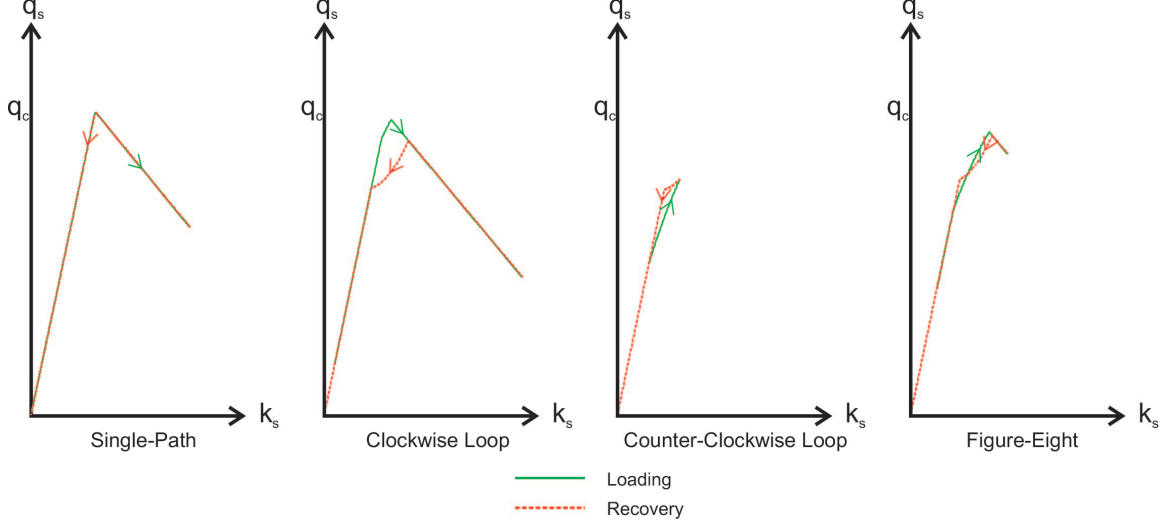


Figure 4.6. Types of patterns observed on the flow-density plane during an LR cycle. Note that the down-sloping ‘tails’ to the right of the loops occur when both bins are simultaneously congested; i.e., when the system is in the CC regime.

(whether at exogenously determined or endogenously determined rates) than as vehicles exit the system. The translation of these paths from the phase diagram to the flow-density plane is now examined to see what types of patterns should be expected due to this asymmetric behavior, and to determine the conditions that define each pattern.

To do this, the evolution of the system during a period of loading followed by a period of recovery (called a loading/recovery or LR cycle) is studied. An LR cycle has two phases: 1) an evenly balanced loading phase with $e_1 = e_2 = E > 0$ and $x_i = 0$, starting at some initial state in the FF regime (k_1^o, k_2^o) until the system reaches a maximum density k_T^* ; and, 2) a recovery phase with $e_i = 0$ and $P_X > 0$ that lasts until the system is completely empty. These LR cycles are used for analysis because they mimic a variety of idealized peak periods by varying only three parameters—the initial starting point (k_1^o, k_2^o) and maximum density k_T^* . Four different types of patterns are discovered: single-path, clockwise hysteresis loops, counter-clockwise hysteresis loops, and figure-eight patterns; see Figure 4.6.

Single-path patterns are named so because they exhibit the same flows for the same density during both the loading and recovery phases, and all points lie on the theoretical MFD of the system. Recall from the description of the phase diagram for the symmetric two-bin system and its translation onto the flow-density plot (Section 3.2.1) that this will only occur if the system remains in the FF and/or CC regimes during the entire LR cycle. This can happen in two ways: either (k_1^o, k_2^o) is perfectly balanced (i.e., $k_1^o = k_2^o$), in which case the system remains perfectly balanced during the entire LR cycle, or else the final state after loading with density k_T^* is in the FF regime so that the phase path during the LR cycle is entirely constrained within the FF regime. Note that the first condition cannot be sustained for long periods

if there is any randomness in turning, entering or exiting flows; minor perturbations will cause the system to move away from a perfectly balanced state. The second case only occurs if the system does not get congested.

Clockwise hysteresis loops exhibit higher flows as the system density increases than as it decreases; i.e., higher flows are observed during loading than during recovery. Since the equi-flow contours in Figure 3.3a show that lower flows are observed as the system becomes more unbalanced, clockwise loops only occur when the system recovers along a path that is further from the diagonal than the path along which it was loaded. This occurs if (k_1^o, k_2^o) is moderately imbalanced so that the path taken during loading approaches the diagonal, and k_T^* is so high that the system enters the D-area and recovers along a path that is more unbalanced than the path taken during loading. These conditions are easy to meet—small disturbances on real networks will cause moderate imbalances during loading and these imbalances will become exacerbated once the network becomes congested. Thus, clockwise hysteresis loops should be very common.

Counter-clockwise hysteresis loops exhibit lower flows as the system density decreases than as it increases; i.e., lower flows are observed during loading than during recovery. Again, since equi-flow contours on the phase diagram show that lower flows are observed as the system becomes more unbalanced, counter-clockwise loops only occur when the system recovers along a path that is closer to the diagonal than the path along which it was loaded. This occurs if (k_1^o, k_2^o) is very unbalanced so that the path taken during loading is far from the diagonal and k_T^* is so low that the system enters the FC regime only briefly without entering the D-area. In this case, based on the phase paths shown in Figures 4.4a and 4.5b, the system will continually move towards a balanced state during the LR cycle, resulting in higher flows during recovery than during loading. These conditions are very specific and can only be met for a small set of initial conditions.

Figure-eight patterns are a combination of clockwise and counter-clockwise hysteresis loops—these patterns exhibit higher flows during loading for high densities and lower during loading for lower densities. This pattern occurs when the path taken during recovery crosses the path taken during loading somewhere in the FC regime. This occurs if (k_1^o, k_2^o) is very unbalanced and k_T^* is very high. These conditions are unlikely as vehicles are not typically unevenly distributed across a network at the beginning of a rush hour period. Note also that the resulting loops in the figure-eight pattern are very small.

4.3.1 Tests with a more realistic traffic network

Since the two-bin model can represent a grid network in which turns are perfectly correlated, it is also of interest to verify if the dynamic behavior predicted by the two-bin model also occurs if turns are not perfectly correlated. An interactive simulation is used to emulate the behavior of vehicles traveling on a more realistic grid network in

which turning maneuvers at each intersection are treated independently from turning maneuvers at other intersections.¹

To confirm the existence of clockwise hysteresis loops during a typical rush hour period, try the following steps with the online simulation: 1) Under the ‘demand’ tab, set the Inflow NS to 8% and press enter (to mimic the loading process); and, 2) once the network becomes congested, set the Inflow NS to 0% and press enter (to mimic the recovery process).

During the loading process, observe on the ‘City MFD’ plot (which shows the average flow on the vertical axis and the average density on the horizontal axis) that flow initially increases with density but then starts to decrease as the network becomes congested; see Figure 4.7a. As the network becomes congested, queue lengths (represented by the black dots on each link) tend to be about the same across all links; see Figure 4.7b.

During the recovery process, observe on the ‘City MFD’ plot that average flows may be initially about the same as those observed during loading. This ‘tail’ occurs because all links are still congested. As the recovery process continues, however, average flows observed during recovery become significantly lower than those observed during loading; see Figure 4.8a. During this time queue lengths are no longer the same across all links—some areas have very long queues (i.e., are congested) while others have very short queues (i.e., are uncongested); see Figure 4.8b. When the density becomes very small, congestion becomes evenly distributed across the network again and points on the plot of average flow vs. average density rejoin those observed during loading.

4.4 Effect of adaptive drivers

Again, the results so far are a ‘worst case’ because they assume drivers follow predetermined and rigid routes so that the rate at which vehicles turn from one bin into the other is a fixed proportion, P_T , of the circulating flow in that bin. Since Section 3.4 shows that the presence of drivers that alter their routes in response to congestion can help smooth the distribution of traffic at equilibrium states, it is possible that the presence of these adaptive drivers can reduce or eliminate the hysteresis loops that are observed when drivers behave rigidly. To study the effect of adaptive drivers on network dynamics systematically, in this section we assume that some proportion, $a \in [0, 1)$, of the drivers behave adaptively and will not turn from a less congested bin to a more congested bin. The rate that vehicles turn from one bin to another is given

¹This simulation was originally created by Professor Jorge Laval of the Georgia Institute of Technology to examine the shapes of macroscopic relationships between flow and density in an open grid network; see <http://trafficlab.ce.gatech.edu/node/19611> and <http://trafficlab.ce.gatech.edu/node/1971> for the original simulation and more details about its operation, respectively. The simulation was modified to make the network closed (so that vehicles leaving a downstream link reappear on an upstream link). This modified simulation is available at <http://www.ce.berkeley.edu/~daganzo/Simulations/classes/CityApplet.html>. The reader is invited to interact with the simulation to verify the phenomena described here.

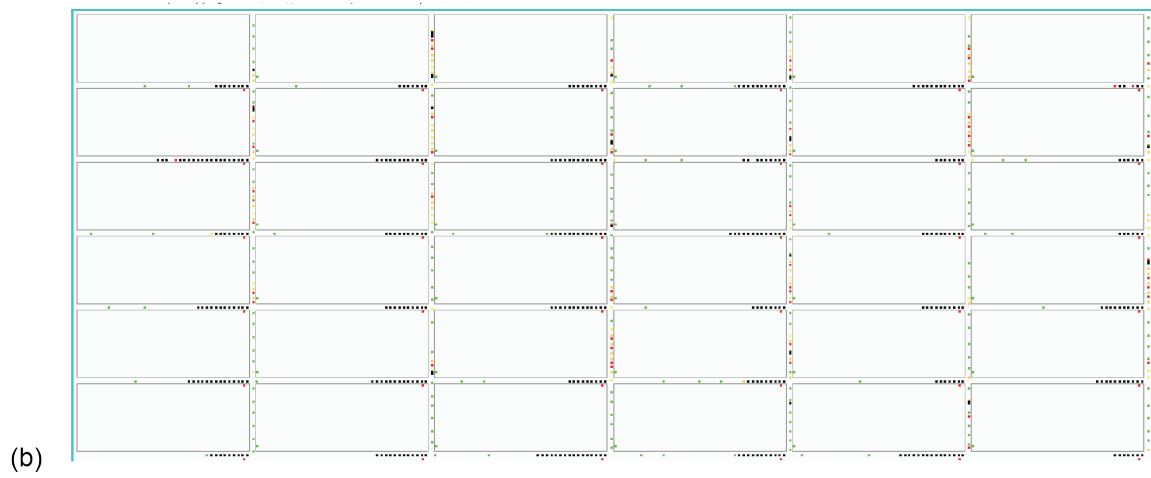
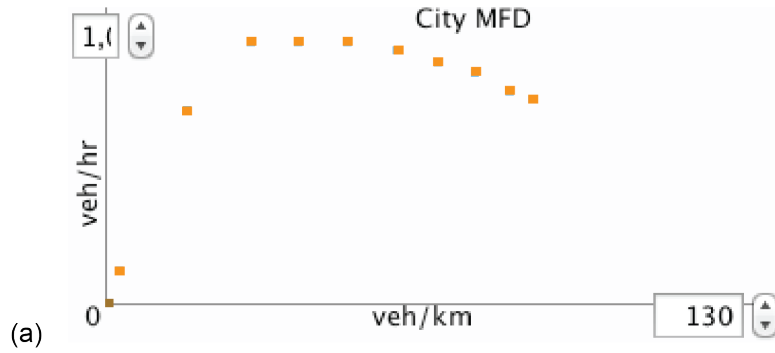


Figure 4.7. More realistic simulation results during loading: (a) CityMFD plot; and (b) congestion pattern across the network.

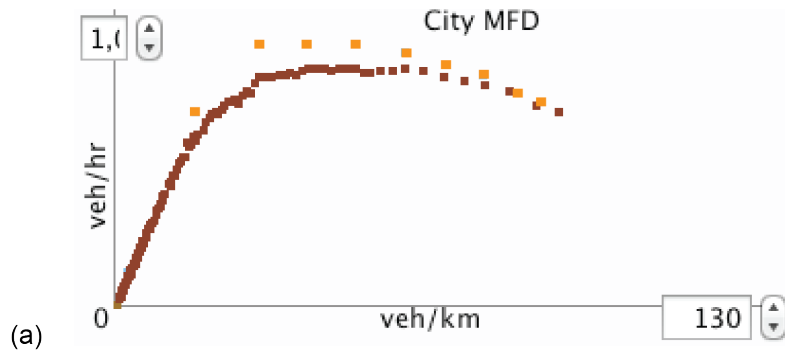


Figure 4.8. More realistic simulation results during recovery: (a) CityMFD plot; and (b) congestion pattern across the network.

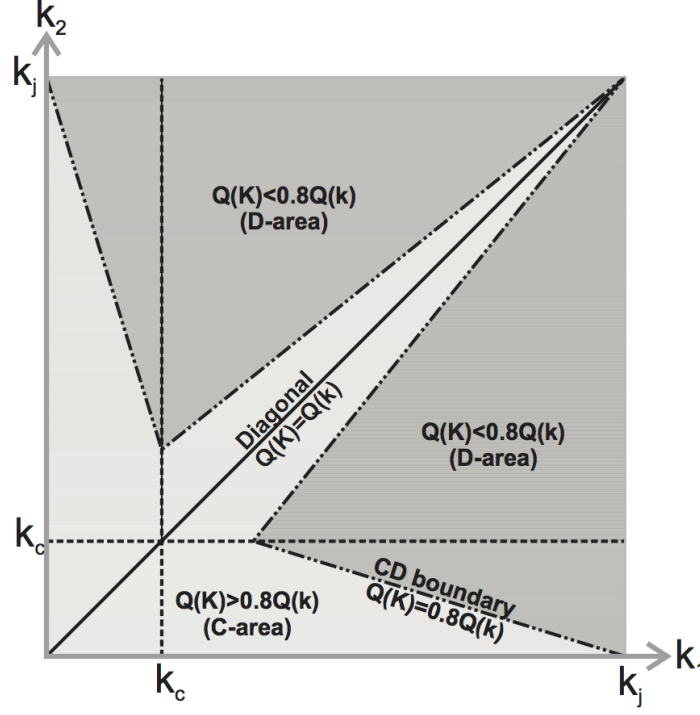


Figure 4.9. Areas of convergence to and divergence from even vehicle distributions when drivers are adaptive for $C_L = 0.8$ (balanced, exogenous loading) and $C_R = 0.8$ (recovery).

by (3.14). Note that a cannot be exactly 1 because drivers near their destination are not able to change their routes and adapt.

With this modification, the dynamic equations for the system during the loading process become:

$$\frac{dk}{dt} = \frac{E + P_T Q(K) - (1 - a) P_T Q(k)}{L} \text{ and } \frac{dK}{dt} = \frac{E + (1 - a) P_T Q(k) - P_T Q(K)}{L}, \quad (4.20)$$

and the convergence measure for loading changes to:

$$\tau_L = \frac{d(k - K)}{dk_T} = \frac{2P_T}{E} [Q(K) - (1 - a)Q(k)]. \quad (4.21)$$

Examination of (4.21) shows that during loading the system converges towards a balanced state when $Q(K) > C_L Q(k)$ and diverges away from it when $Q(K) < C_L Q(k)$, where $C_L = 1 - a$. Notice that the C-area grows and the D-area shrinks as a increases and drivers become more adaptive. The C-area and D-area for the case in which $a = 0.8$ are plotted in Figure 4.9. Also notice that (4.21) is monotonically increasing with respect to a ; therefore, the tendency towards evenness increases with the adaptivity of drivers.

The dynamic equations during the recovery process when drivers are adaptive are:

$$\frac{dk}{dt} = \frac{pQ(K) - ((1-a)P_T + P_X)Q(k)}{L} \text{ and } \frac{dK}{dt} = \frac{(1-a)pQ(k) - (P_T + P_X)Q(K)}{L}, \quad (4.22)$$

and the convergence measure changes to:

$$\tau_R = -\frac{d(k-K)}{dk_T} = \frac{1}{P_X q_S} [(2P_T + P_X)Q(K) - (2(1-a)P_T + P_X)Q(k)]. \quad (4.23)$$

Examination of (4.23) shows that during recovery the system converges toward a balanced state when $Q(K) > C_R Q(k)$ and diverges away from a balanced state when $Q(K) < C_R Q(k)$, where $C_R = 1 - 2aP_T/(2P_T + P_X)$. Note that if $C_L = C_R$ the C- and D-areas are the same for both loading and recovery. Thus, Figure 4.9 also presents the C- and D-areas for $C_R = 0.8$. As with the loading process, the C-area grows and the D-area shrinks as drivers become more adaptive. And since (4.23) is monotonically increasing with respect to a , the tendency toward evenness continues to increase with the adaptivity of the drivers. For any given value of a , $C_L \leq C_R$; therefore, the C-area is larger during loading than during recovery. Consideration of this fact and a comparison of (4.21) and (4.23) shows that the system tends to evenness for a wider range of states and does so more strongly during loading than during recovery.

Sample loading paths are presented in Figure 4.10 for the loading and recovery processes with adaptive drivers. Unlike the non-adaptive case, paths that start at an unbalanced state are now able to reach the diagonal. Thus, it is possible for a single-path pattern that lies along the theoretical MFD of the system to occur during an LR cycle even when the system starts out of balance and becomes very congested. The range of starting states that converge to the diagonal during loading and recovery increases with the adaptiveness of the drivers. Therefore, the probability of observing the single-path regime increases (and the probability of observing clockwise hysteresis loops decreases) with the adaptiveness of the drivers.

This behavior is also confirmed when the rate at which vehicles turn between the two bins is allowed to fluctuate randomly around $P_T Q(k_i)$. In this case, the starting state (k_1^o, k_2^o) becomes less important because random variations in the turning flows cause the random disturbances in the system that change the balance of vehicles. Some sample phase paths are presented in Figure 4.11 for two levels of driver adaptivity when the system begins and ends completely empty. Note that when drivers are less adaptive (Figure 4.11a for which $a = 0.3$) small imbalances due to randomness can grow with time. If this happens, then clockwise hysteresis loops will arise on the flow-density plane. However, when drivers are very adaptive (Figure 4.11b for which $a = 0.7$) imbalances in the system tend not to grow. Phase paths stay close to the diagonal during both loading and recovery and single-path patterns are observed consistently on the flow-density plane.

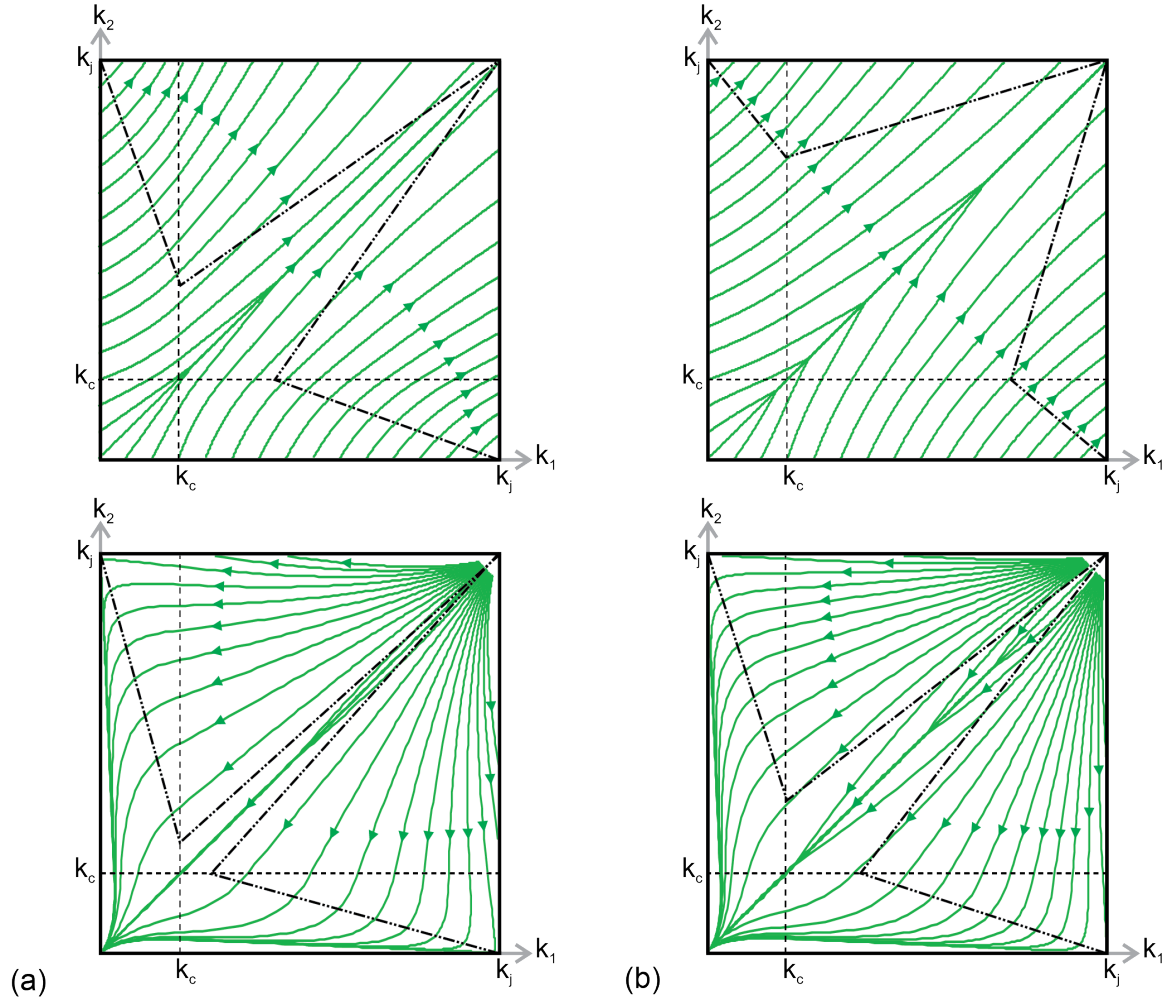


Figure 4.10. Balanced, exogenous loading paths (top) and recovery paths (bottom) for two adaptation levels: (a) $a = 0.3$; and, (b) $a = 0.7$. In all cases, $E = 0.2q_c$, $P_X = 0.2$ and $P_T = 0.05$.

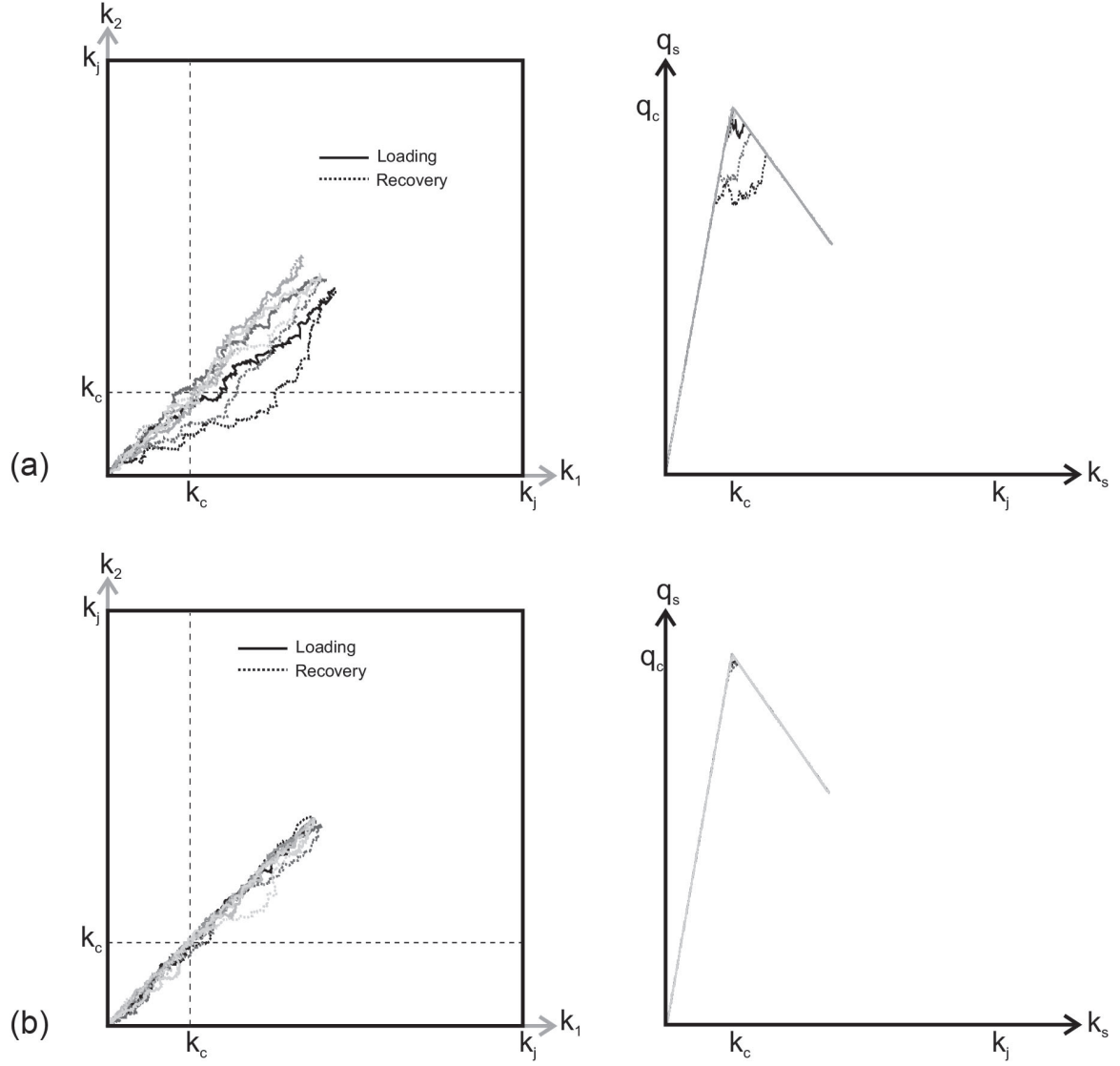


Figure 4.11. Phase paths with stochastic turning, $E = 0.2q_c$, $P_T = 0.05$, $P_X = 0.2$, loading period of 500 time-steps of $\Delta t = 1/1200$ hr, $(k_1^o, k_2^o) = (0, 0)$ and: (a) $a = 0.3$; and, (b) $a = 0.7$. Note what appears to be portions of a triangle on the flow-density plot are actually sample paths in the CC regime.

4.5 Major findings and implications for more realistic networks

Studying the dynamics of the two-bin model shows that networks are inherently more unstable as the average density decreases than as the average density increases. This occurs even when turning is non-existent. The reason is that during recovery more congested areas clear more slowly than less congested areas because the rate that vehicles exit a bin is a fixed proportion of the flow within the bin. During loading vehicles either enter more congested and less congested areas at the same rate (exogenous loading) or enter the more congested areas at a lower rate (endogenous loading). Thus, in the absence of turning, imbalances in the system will either always remain fixed or decrease with time during loading, whereas these imbalances will tend to increase with time as the network recovers from congestion.

The presence of turning exacerbates the natural instability present during recovery, and introduces an instability during loading. This instability causes clockwise hysteresis loops to likely occur on the aggregate flow-density plot for the system. Other factors, such as certain spatio-temporal distributions of origins and destinations, can also contribute to the existence of hysteresis (both clockwise and counter-clockwise) in the aggregate flow-density relationship of traffic networks. However, this idealized model shows that clockwise loops should be expected even in the most favorable conditions of perfect network symmetry and perfectly balanced demand. The instability caused by turning is found to decrease if drivers adaptively re-route themselves in real-time to avoid congested areas of the network.

The qualitative behavior of the two-bin system studied in this chapter can be extended to more realistic traffic networks since real networks can be modeled as a similar system of many more interconnected bins. The qualitative behavior of these n -bin systems should be the same as the qualitative behavior of the two-bin system analyzed here. Although it is very unlikely that all n bins would be congested simultaneously, we still expect two effects: (i) recovery to be more unstable than loading (because more congested areas will clear more slowly than less congested areas); and, (ii) this instability to increase with the lack of driver adaption.

Figure 4.12 presents empirical data that seem to validate the qualitative behavior predicted by modeling a traffic network as a system of bins. Figures 4.12a and 4.12b are the aggregate flow-density relationship of Yokohama, Japan and Toulouse, France obtained from empirical observation. Both were obtained over the course of a rush hour period and contain data taken from when network densities were increasing and decreasing. Notice that all observations from the Yokohama dataset fall approximately along the same curve; a single value of flow is observed consistently for each density. This makes sense based on the results of the two-bin model—real drivers are highly adaptive and this high level of adaptivity should result in single-path patterns in the aggregate flow-density relationship.

The Toulouse observations follow the same pattern except for the data from one particular day (June 13, 2008), which are magnified with time stamps in the inset of

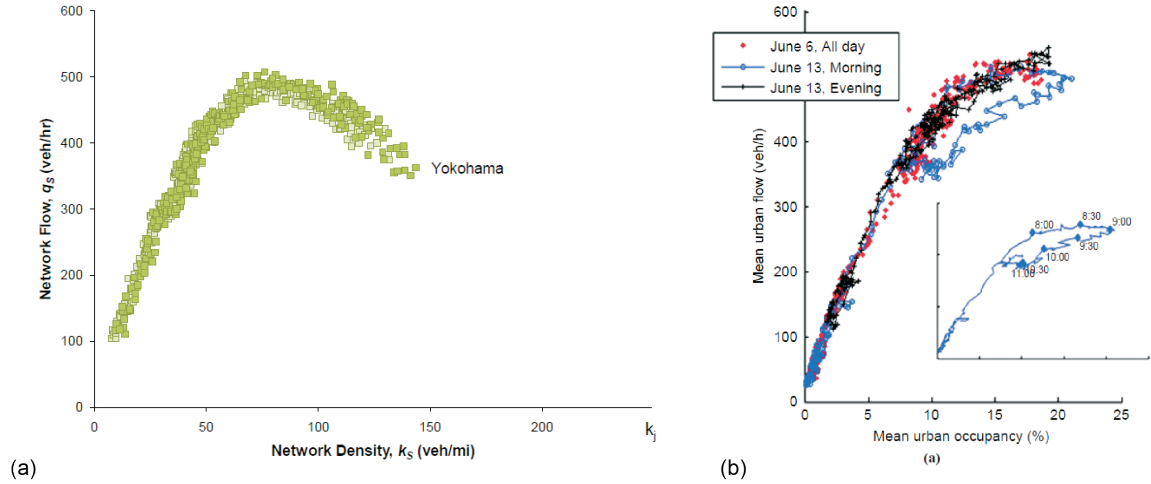


Figure 4.12. Empirical MFDs of: (a) Yokohama, Japan (source: Geroliminis & Daganzo, 2008); and, Toulouse, France (source: Buisson & Ladier, 2009).

Figure 4.12b. On this day, there appears to be a clockwise hysteresis loop with higher flows during the beginning of the rush and lower flows during the end of the rush. Based on the results of the two-bin model, such a pattern would only be expected if some sort of (large) disturbance occurred that caused traffic to become very unevenly distributed in the network, and if drivers were not adaptive enough to compensate for this unevenness and it was allowed to grow with time. Digging deeper, it was noted in Buisson & Ladier (2009) that on this particular day a social movement of truckers caused reduced average speeds and created unusually congestion on the ring road surrounding the city between 8:00 AM and 10:00 AM. This caused drivers to change routes and flee for the surrounding urban streets. This would have created an unusual imbalance in the congestion distribution across the urban network. Additionally, it can be imagined that drivers on the city streets would become less adaptive, since some would have switched off the ring road to unfamiliar routes, and this would result in a clockwise hysteresis loop. As in the two-bin simulations, the low flows were mainly observed during recovery and the hysteresis pattern persisted until the end of the rush. Effect (i) is perhaps why low flows were observed only when the network was recovering from congestion, and effect (ii) is perhaps why this happened on the day of the strike. The combination of effects (i) and (ii) also help to explain why these low flows continued after the truck strike ends.

The results of the two-bin model also suggest that clockwise hysteresis loops should arise consistently in the aggregate flow-density relationship of freeway networks. The topology of these systems is such that drivers have fewer opportunities to adapt (either by exiting the system and moving to a surface street, or switching between routes in the network). Indeed, recent empirical data confirms the existence of clockwise hysteresis loops on freeway systems (Geroliminis & Sun, 2011a,b).

Chapter 5

Effect of turning on network efficiency

Chapters 2, 3 and 4 showed that the presence of turning between multiple routes creates an instability in urban traffic networks, and this instability causes chaotic aggregate behavior. Fortunately, Chapter 3 also showed that if drivers adapt to avoid congestion this instability can be mitigated and maximum flow sustained for long periods. Chapter 4 also showed that the presence of adaptive drivers can balance network behavior during the beginning and end of a rush hour, so that maximum flows can be achieved both as congestion forms and dissipates. Therefore, a logical next step is to see what these maximum flows are and how they might be affected by turning between routes.

It is well known that conflicting left turns can reduce flows at an intersection. For example, left-turning vehicles served in a permitted fashion must wait for a gap in the opposing traffic and can block upstream vehicles during this time (Newell, 1959; Fambro *et al.*, 1977). Separate lanes can be used to segregate left-turning vehicles from other vehicles to reduce this blocking effect. However, this strategy does not eliminate blocking if the lane is not long enough to accommodate left-turn queues or if left-turn lanes are blocked by queues on adjacent lanes. This strategy also reduces the amount of space available for other vehicles to discharge (Messer & Fambro, 1977; Wong & Wong, 2003). Protected phases can also be used to eliminate blocking (Cottrell, 1986), but this increases the amount of time lost for movement during a cycle. Since maximum network flows are limited by maximum flows through the intersections (Daganzo & Geroliminis, 2008), lower flows at the intersection result in lower maximum network flows. For this reason, conflicting left turns are sometimes eliminated by using one-way streets.

However, maximum network flows might not be the best metric with which to compare different treatments of conflicting left turns. Daganzo (2007) shows that during peak periods the optimal control strategy for a network is to limit vehicle entry so that the trip completion rate is always maximized. Such a strategy minimizes the total travel time of all vehicles using the network. This work also suggests

that network performance can be improved if the maximum trip completion rate is increased. Therefore, the maximum trip completion rate (and not maximum network flow) might be a better metric to compare networks, and compare different treatments of conflicting left turns. Eliminating conflicting left turns by using one-way streets might actually make network performance worse if it is associated with a lower maximum trip completion rate, even if maximum flows are higher.

In view of the above, this chapter presents an analytical model that can be used to estimate the maximum trip completion rates (or trip-serving capacities) of different networks based on their treatment of conflicting left-turning maneuvers. The insights obtained from this model can be used to determine conditions under which allowing conflicting turns is beneficial and identify when these movements should be eliminated. Section 5.1 describes the capacity formula that is used, which gives compares the maximum trip-serving rate of a network with an idealized network. Section 5.2 describes the different networks considered based on their treatment of left turns. Section 5.3 shows how the parameters of capacity equation can be obtained from a few known network properties. Section 5.4 then presents a comparison of the maximum efficiencies of the different networks, and unveils conditions under which conflicting left-turning maneuvers should be allowed to maximize network performance. Section 5.5 discusses qualitatively how the results presented here might change on more realistic networks. Finally, Section 5.6 summarizes the major findings of this chapter and implications for recent trends in network operations.

5.1 Capacity formula

Considered here is a homogeneous grid network with long blocks, similar to the one pictured in Figure 2.1, except streets are not necessarily assumed to be one-directional. It is assumed that origins and destinations are evenly spread across the network so that travel is evenly split between the two families of streets, and that the average trip length is known and stable over time.

The analysis starts with the operation of the intersections because both the vehicle-moving capacity and trip-serving capacity of a street network of the type analyzed here is mainly constrained by the capacity of individual intersections if blocks are long (Daganzo & Geroliminis, 2008). It is assumed that all intersections are signalized, that available green time is evenly split between serving the two families of streets, and that the lost time between consecutive phases is constant. It is also assumed that left turns are only allowed during separate, protected left-turn phases (if possible) and that these left-turn phases are perfectly timed to serve the left-turn demand with no wasted time. This last assumption is optimistic but will provide an upper bound on the trip-serving capacities of networks that allow conflicting left turns. The effects of relaxing this and other simplifying assumptions will be discussed in Section 5.5.

The analysis starts by envisioning a network made up of two-way streets with intersections that have wide lanes and overpasses with ramps to facilitate turning

movements, so that conflicting movements are physically separated; see Figure 5.1. Furthermore, it is assumed that travel time on ramps is negligible. Since the number of vehicles able to move through an intersection is proportional to the product of the amount of green time available and the number of lanes available for discharge, measured in lane-seconds, an intersection operated as in Figure 5.1 should be able to serve the highest vehicle flow because vehicles are able to discharge continuously and do so using all lanes. Additionally, since two-way streets are used and all turn movements are allowed, vehicles are able to take the most direct routes to reach their destinations. This allows trips to be completed at the highest rate possible. More realistic networks with signalized intersections instead of overpasses should have lower vehicle-moving capacities for three main reasons: 1) red phases at the intersection signals would prohibit vehicles from discharging during a portion of the cycle; 2) physical separation of through and left-turn movements would mean fewer lanes are able to discharge at once; and, 3) narrower lanes used to accommodate left-turn pockets would prohibit vehicles from discharging at the maximum rate. Additionally, trip-serving capacities may be lower on more realistic networks because one-way streets or turn restrictions (if implemented) would force vehicles to travel longer distances on average, which results in a lower trip completion rate (Daganzo, 2007).

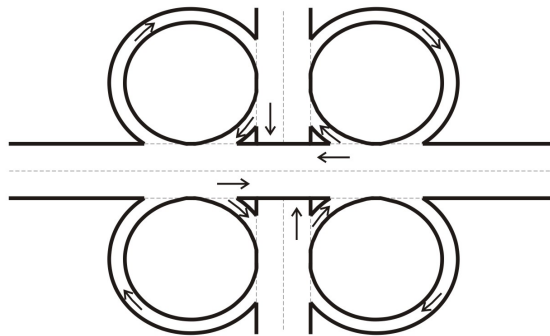


Figure 5.1. Intersection configuration for idealized network in capacity analysis.

The idealized unsignalized network just described is used as a benchmark against which to compare other signalized networks that treat conflicting left turns differently. The relative trip-serving capacity of each network, denoted C_j , is given as the proportion of the maximum rate at which trips can be made in network j as compared to the maximum rate at which trips can be made in the idealized network.¹ For the remainder of this chapter, the term capacity will refer to the trip-serving capacity defined in this way. This capacity can be expressed as the product of: (i) the ratio of lane-seconds of discharge time available in network j compared to the idealized network; (ii) the ratio of saturation flows through the intersection in network j compared to the idealized network; and, (iii) the inverse of the ratio of average trip lengths in

¹For the remainder of this Chapter, the term capacity will refer to the trip-serving capacity defined in this way.

network j compared to the idealized network. Since the two families of streets are symmetric and treated the same way, the capacity of network j can be determined by examining the operation of an intersection as it serves one of these directions, as follows:

$$C_j = \left[\frac{N_j^T G_j^T + N_j^L G_j^L}{N_j C} \right] \frac{s_j}{\alpha_j} = \left[\frac{N_j^T g_j^T + N_j^L g_j^L}{N_j} \right] \frac{s_j}{\alpha_j}, \quad (5.1)$$

where N_j^T and N_j^L are the total number of through and left-turn lanes, respectively, available for vehicles to discharge on one family of streets for network j , N_j the total number of wide lanes that would be available for one street family in an idealized network using the same space, G_j^T and G_j^L the green time available per cycle for through and left-turning vehicles, respectively, to discharge on one street family, C the length of the signal cycle, s_j the ratio of saturation flows at the intersection in network j compared to the idealized network, and α_j the ratio of average trip lengths in network j compared to the idealized network.²

The numerator of the term in brackets in the first equality of (5.1) is the lane-seconds of green available per cycle for both through-moving and left-turning vehicles traveling on one family of streets in network j , and the denominator is the lane-seconds that would be available for the movement of vehicles on one family of streets in the idealized network. Therefore, the term in brackets is the ratio of lane-seconds of green available for discharge in network j compared to the idealized network. This ratio multiplied by the ratio of saturation flows and the inverse of the ratio of average trip lengths yields the fraction of trips possible per unit time using network j as compared to the idealized network. This capacity can be equivalently expressed as a function of the fraction of green time per cycle available for through and left-turn movements (denoted g_j^T and g_j^L , respectively) as shown in the second equality of (5.1).

5.2 Networks considered

This section describes the different networks considered based on how conflicting left-turning maneuvers are treated. The intersection configurations for the networks with different conflicting left-turn treatments that are to be considered are shown in Figure 5.2. The networks are: a two-way network with overlapping left-turn lanes ($j = a$); a two-way network with non-overlapping left-turn lanes ($j = b$); a two-way network with left-turn pockets ($j = c$); a two-way network with pre-signals ($j = d$); a two-way network with single lane intersection approaches ($j = e$); a two-way network with banned left-turns ($j = f$); and, a one-way network ($j = g$). Note that only networks that are symmetric and have the same treatments on all links are considered.

²Note that the product of (i) and (ii) yields the relative vehicle-moving capacity of network j as the proportion of the maximum rate vehicles can move in network j as compared to the idealized network. Thus, (5.1) can be used to compare vehicle-moving capacities if the α_j term is removed.

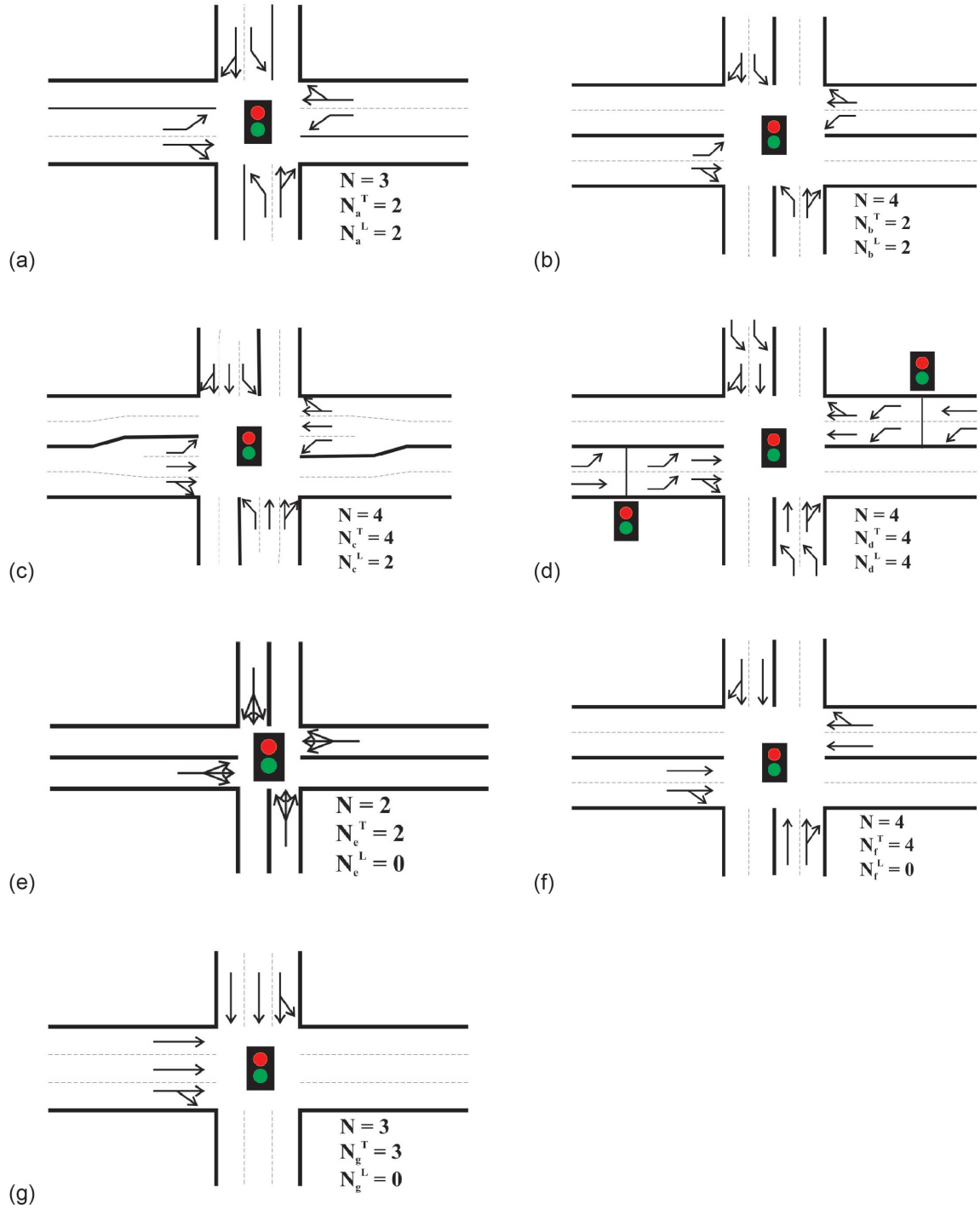


Figure 5.2. Intersection configurations for: (a) a two-way network with overlapping left-turn lanes; (b) a two-way network with non-overlapping left-turn lanes; (c) a two-way network with left-turn pockets; (d) a two-way network with pre-signals; (e) a two-way network with single-lane approaches; (f) a two-way network with banned left-turns; and, (g) a one-way network.

Conflicting left turns can be physically separated from through traffic by dedicating a lane for left-turning maneuvers at the intersection. These lanes can be overlapping or non-overlapping, as shown in Figure 5.2a and 5.2b, respectively. In the case where these left-turn lanes overlap, only one lane of space is required to serve left-turning vehicles approaching from opposite directions. However, when left-turn lanes do not overlap, two lanes of space are taken from through vehicles to serve the left turns at opposing approaches. Intuitively, the non-overlapping case is less efficient than the overlapping case, so overlapping turn lanes should be used whenever possible. However, networks with overlapping left-turn lanes can only be used when there are an odd number of lanes available at the intersection if symmetry is to be maintained.

Taking a lane away from through-moving vehicles is not the only way to serve left-turning vehicles. In some cases, lanes can be narrowed upstream of the intersection to create additional space, called a turn pocket, that left-turning vehicles can use to discharge at the intersection; see Figure 5.2c. This provides more lane-seconds of green for vehicles to discharge, but reduces the discharge rate of these vehicles through the intersection because they must now navigate through narrower lanes.

Left-turning vehicles can also be segregated longitudinally along the roadway length, instead of laterally as done with dedicated lanes. Additional signals (called pre-signals) can be installed upstream of the intersection approach to re-organize vehicles at the intersection in such a way that through-moving and left-turning vehicles are able to discharge from all lanes during their respective phases (Xuan *et al.*, 2011, 2012); see Figure 5.2d.³ The pre-signal strategy requires long blocks to accommodate queues at the mid-block signal; however, since networks with long blocks are already being considered here this should not present an additional problem.

Figure 5.2e shows the special case of a network that allows left turns, but forces them to be mixed with through vehicles. This can occur when each intersection approach is served by only a single lane. In this case, protected left-turn phases cannot be used at the intersection signal because left-turning and through-moving vehicles are mixed together. Instead, both must discharge during the same phase. Left-turning vehicles must wait for a gap in the opposing traffic (which could be produced by the arrival of another left-turning vehicle) before proceeding. During this time the waiting left-turner will block all vehicles behind it. This blocking reduces the maximum average discharge rate through the intersection.

Another way to treat left turns is to simply ban them altogether as shown in Figure 5.2f. This movement restriction will reduce the number of phases required at the intersection. Thus, less lost time will be incurred and more lane-seconds will be available for discharge. However, banning these movements will also increase average trip lengths as vehicles will be forced to take more circuitous routes. Similarly, one-way streets could be used throughout the network as in Figure 5.2g. This would eliminate the potential for conflicting turning movements and result in simpler phasing schemes, but will also increase average trip lengths.

³Note that a similar idea has also been proposed to segregate different modes (Xuan *et al.*, 2010).

5.3 Network parameters

The parameters of (5.1), and therefore the capacities of the different networks, turn out to depend only on a few key characteristics: the number, type and width of the lanes available for discharge at the intersections, the lost time per phase change and the average trip length in the network.

5.3.1 Number and type of lanes (N_j^T, N_j^L, N_j)

The number and type of lanes available to serve vehicles on one family of streets for each network are easily determined by examining the geometry of the intersections. Values for each network are shown in Figure 5.2 for the specific intersection configurations shown.

For example, in Figure 5.2 network *a* has three lanes of space to serve a particular street family. Assuming these lanes are wide, an idealized network using the same amount of space would have the same number of lanes, so $N_a = 3$. Examining the east-west street family, through-moving vehicles are able to discharge from two lanes: one from east to west and one from west to east, so $N_a^T = 2$. Likewise, left-turning vehicles are able to discharge from two lanes, so $N_a^L = 2$.

In network *c*, four wide lanes are narrowed to provide five narrow lanes of space at the intersection. An idealized network using the same amount of space would only be able to have four lanes, so $N_c = 4$. On the east-west street family, through-moving vehicles are able to discharge from four narrow lanes and left-turning vehicles from only two narrow lanes, so $N_c^T = 4$ and $N_c^L = 2$.

In network *g*, three lanes of space are available to serve each direction, so $N_g = 3$. Since all lanes serve through-moving vehicles and no left-turn lanes (or left-turning movements) exist at the intersection, $N_g^T = 3$ and $N_g^L = 0$.

5.3.2 Green fractions (g_j^T, g_j^L)

For the first four networks presented in Figure 5.2, two phases are needed to serve vehicles on each street family: a protected left-turn phase and a through phase. Since signals are evenly balanced between the two families of streets, half of the cycle length is dedicated to serve each. However, not all of this time can be used by vehicles to discharge because some time is lost between phase changes. If a four-phase signal scheme is used that provides a left-turn and through phase for each family of streets, four lost times will be incurred per cycle. Half of these lost times can be assigned to each street family. Thus, the fraction of cycle available for through and left-turn phases on one street family is:

$$g_j^T + g_j^L = \frac{1}{2} - 2l, \quad i = a, b, c, d; \quad (5.2)$$

where $l \ll 1$ is the fraction of the cycle lost per phase change.

To calculate how this green time can be optimally assigned to serve the through and left-turn phases, assume that the average trip length is n (measured in block lengths between the origin and destination). The average number of left-turns made per trip can be expressed as a function of the average trip length, $L(n)$. The ratio $L(n)/n$ is the average number of left turns made per vehicle-block traveled or, equivalently, the proportion of vehicles that turn left at any intersection, $P_T(n)$. If the signal is perfectly timed so that through and left-turn movements are on the verge of over-saturation, the ratio of lane-seconds green available to left-turning vehicles to that available for all vehicles must equal the proportion of left turns at the intersection; i.e.:

$$\frac{g_j^L N_j^L}{g_j^T N_j^T + g_j^L N_j^L} = \frac{L(n)}{N} \doteq P_T(n), \quad i = a, b, c, d; \quad (5.3)$$

or equivalently:

$$\frac{g_j^T}{g_j^L} = \frac{N_j^T}{N_j^L} \frac{L(n)}{n - L(n)}, \quad i = a, b, c, d. \quad (5.4)$$

The quantity $L(n)$ can be determined based on the routing strategy vehicles use. Assume now that vehicles use the most direct routes between all origins and destinations, and select from among the many shortest paths available by choosing the one that minimizes the number of left turns. For this strategy, probability theory shows:

$$L(n) = \frac{6n - 2}{8n}, \quad n \geq 2. \quad (5.5)$$

Substituting (5.5) into (5.4), we see that the proportion of time needed for the left-turn phase decreases with n and N_j^L but increases with N_j^T . The capacity of networks a, b, c and d can be found by substituting (5.2), (5.4) and (5.5) into (5.1). The terms of this expression can then be simplified, and this yields:

$$C_j = \frac{1/2 - 2l}{N_j} \left[\frac{N_j^T N_j^L n}{N_j^T L(n) + N_j^L [n - L(n)]} \right] \frac{s_j}{\alpha_j}, \quad i = a, b, c, d. \quad (5.6)$$

For the last three networks in Figure 5.2, left-turn phases and left-turn lanes are not needed at the intersection. Therefore, the intersection can be served with a two-phase traffic signal and only one lost time is assigned to each street family. Also, all lanes are available for through-moving vehicles. Therefore, the capacity of these networks simplifies to:

$$C_j = \left[\frac{1}{2} - l \right] \frac{s_j}{\alpha_j}, \quad i = e, f, g. \quad (5.7)$$

5.3.3 Saturation flow ratios (s_j)

For networks a, b, d, f , and g vehicles are assumed to discharge from the intersection at the same rate as the idealized network using the same space; therefore, $s_j = 1$ for $j = a, b, d, f, g$.

For network c , vehicles should discharge at a lower rate since lanes are narrowed at the intersection. The rate at which vehicles discharge through narrow lanes can be determined using the methodology suggested in the Highway Capacity Manual (HCM, 2000). In this case:

$$s_c = 1 + \frac{w_l - 12}{30}, \quad (5.8)$$

where $w_l \leq 12$ is the width of the narrow lanes at the intersection (in feet). For example, if in Figure 5.2c four 12 ft wide lanes are narrowed to provide five 9.6 ft wide lanes at the intersection, the saturation flow fraction calculated from (5.8) is $s_c = 1 + (9.6 - 12)/30 = 0.92$.

For network e , left-turning vehicles block through vehicles at the intersection and this reduces the average rate at which vehicles can discharge. This average saturation flow ratio can be found by treating vehicles discharging through the intersection as a Markovian process. The regeneration point of this Markov chain is the presence of two opposing vehicles to the intersection, and each regeneration period is indexed by m . Assume that vehicles are queued at each of the intersection approaches so that the regeneration period occurs consistently every saturation headway, $1/q_s$, where q_s is the saturation flow of vehicles in the idealized network without narrow lanes or blocking. The state of the Markov chain is the number of vehicles that are able to discharge through the intersection between regeneration points.

If opposing vehicles at the intersection are making the same movement (e.g., both are left-turning or through-moving) then both will be able to discharge and $x_m = 2$. However, if the opposing vehicles are of different movements only one will be able to discharge and $x_m = 1$. Define the probability that the left-turning vehicle discharges in this situation as Q_T . The transition probability matrix can be described as a function of $P_T(n)$ and Q_T :

$$\begin{array}{cc} & \begin{array}{cc} x_{m+1} = 1 & x_{m+1} = 2 \end{array} \\ \begin{array}{c} x_m = 1 \\ x_m = 2 \end{array} & \left[\begin{array}{cc} Q_T P_T(n) + (1 - Q_T)(1 - P_T(n)) & Q_T(1 - P_T(n)) + (1 - Q_T)P_T(n) \\ 2P_T(n)(1 - P_T(n)) & P_T(n)^2 + (1 - P_T(n))^2 \end{array} \right] \end{array} \quad (5.9)$$

Analyzing the equilibrium behavior of the Markov chain described by (5.9) shows that the average saturation flow in this case is:

$$\left[\frac{2P_T(n) + Q_T - P_T(n)^2 - 2P_T Q_T}{3P_T(n) + Q_T - 2P_T(n)^2 - 2P_T(n)Q_T} \right] q_s. \quad (5.10)$$

For the special case in which left-turning vehicles must yield to through-moving

vehicles, which occurs when left-turning vehicles must wait in a gap in the opposing traffic stream, $Q_T = 0$, and the saturation flow fraction simplifies to:

$$s_e = \frac{2 - P_T(n)}{3 - 2P_T(n)}. \quad (5.11)$$

Equation (5.11) shows that the saturation fraction, s_e , decreases as $P_T(n)$ decreases, approaching $2/3$ as $P_T(n)$ approaches 0. This makes physical sense: as left turns become more infrequent, any left-turning vehicle that approaches the intersection must wait at the intersection (blocking all vehicles behind it) for a longer time on average until an opposing left-turning vehicle arrives and they can both move. Since $P_T(n) = L(n)/n$, (5.11) shows that the saturation fraction decreases with n for this network.

5.3.4 Average trip length ratios (α_j)

For all networks except networks f and g , vehicles are able to use the most direct route to reach their destination because there are no restrictions on movement through the network; thus $\alpha_j = 1$ for $j = a, b, c, d, e$.

For networks f and g , however, vehicles will have to travel longer routes on average to reach their destination either due to the left-turn or one-way restrictions. Probability theory was used to determine the average additional distance traveled (in block lengths) from an origin to any destination when these restrictions are in place. The additional travel distance for networks f and g are both found to be a function of the average trip length and are denoted $D_f(n)$ and $D_g(n)$, respectively. The additional distance on two-way networks with banned left turns, $D_f(n)$, is found to increase with n and approaches an upper bound of 1 additional block length traveled per trip. The additional travel distance on one-way networks, $D_g(n)$, is found to generally decrease with n and approaches a lower bound of 2 additional block lengths traveled per trip.⁴ Thus, the ratio of average trip lengths in a two-way network with banned left-turns as compared to the idealized network is:

$$\alpha_f(n) = \frac{n + D_f(n)}{n} < 1 + 1/n, \quad (5.12)$$

and the same ratio for the one-way network is:

$$\alpha_g(n) = \frac{n + D_g(n)}{n} > 1 + 2/n. \quad (5.13)$$

Since (5.12) and (5.13) imply that $\alpha_f(n) > \alpha_g(n)$ for all n , this means that one-way networks impose more circuitous routes than two-way networks with banned left

⁴ $D_g(n)$ is said to generally decrease with n because it is not a strictly decreasing function. Due to the geometry involved with one-way networks, $D_g(n)$ strictly decreases for increasing even values of n and also strictly decreases for increasing odd values of n . However, if n is even then $D_g(n+1)$ is slightly greater than $D_g(n)$ for $n \geq 8$ due to the geometry of one-way street networks.

turns. This makes physical sense since one-way networks are more restrictive and prohibit more movements. In both cases, the proportion of additional travel distance decreases with n ; therefore, as trips become longer the penalty imposed due to the circuitous routing decreases.⁵

5.4 Network comparison

Equations (5.6), (5.7), (5.8), (5.11), (5.12) and (5.13) are now combined and applied to evaluate the capacities of the networks with different conflicting left-turn movement treatments. Since g_j^T , g_j^L , s_j and α_j are functions of n and l , the capacities of each network are plotted vs. n for two different values of l . Figures 5.3, 5.4 and 5.5 present the capacities of different two-, three- and four-lane networks, respectively, for $l = 0.00$ and $l = 0.05$. Note that only the treatments associated with networks e , f and g can be applied to two-lane networks. Similarly, only networks a and g be applied to three-lane networks and only networks b , c , d , f and g can be applied to four-lane networks.

The capacities of networks a , b and e do not vary with n as shown in Figures 5.3, 5.4 and 5.5. This occurs because in these networks vehicles use the same number of lanes to discharge during both the left-turn and through phases (two, two and four, respectively, for these network types). Thus, even though the lengths of the left-turn and through phases (g_j^T and g_j^L) vary with n , the total lane-seconds of green available for vehicles to discharge does not change with the exact split between the green phases or the average trip length. Compare this to network c . In this network, vehicles discharge from only two lanes during the left-turn phase but discharge from four lanes during the through phase. Therefore, the capacity of the network will increase as the left-turn phase becomes shorter and the through phase becomes longer (i.e., as n increases). Also note that higher values of lost time reduce the capacities of networks that use a four-phase signal scheme (networks a , b , c and d) more than networks that use a two-phase signal scheme (e , f and g). This is as expected since more phases result in more lost time incurred.

Figures 5.3, 5.4 and 5.5 show that the treatment of conflicting left-turning maneuvers can significantly affect the capacities of a network. For given values of n , l , and N_j , the maximum rate trips can be served can vary as much as 50% depending on the conflicting left-turn treatment. What is interesting is that using one-way streets to eliminate these conflicting turning maneuvers does not always improve network performance, which contradicts what is suggested by conventional wisdom and current design handbooks. Even though maximum vehicle flows might be higher on one-way networks, the additional circuitry they impose result in a lower maximum trip completion rate, which has been found to be a much better predictor of network performance. Figure 5.6 demonstrates this by comparing the maximum vehicle flows and trip-serving capacities of a four-lane network of one-way streets with a network

⁵This is true for one-way networks even though $D_g(n)$ is not a strictly decreasing function of n .

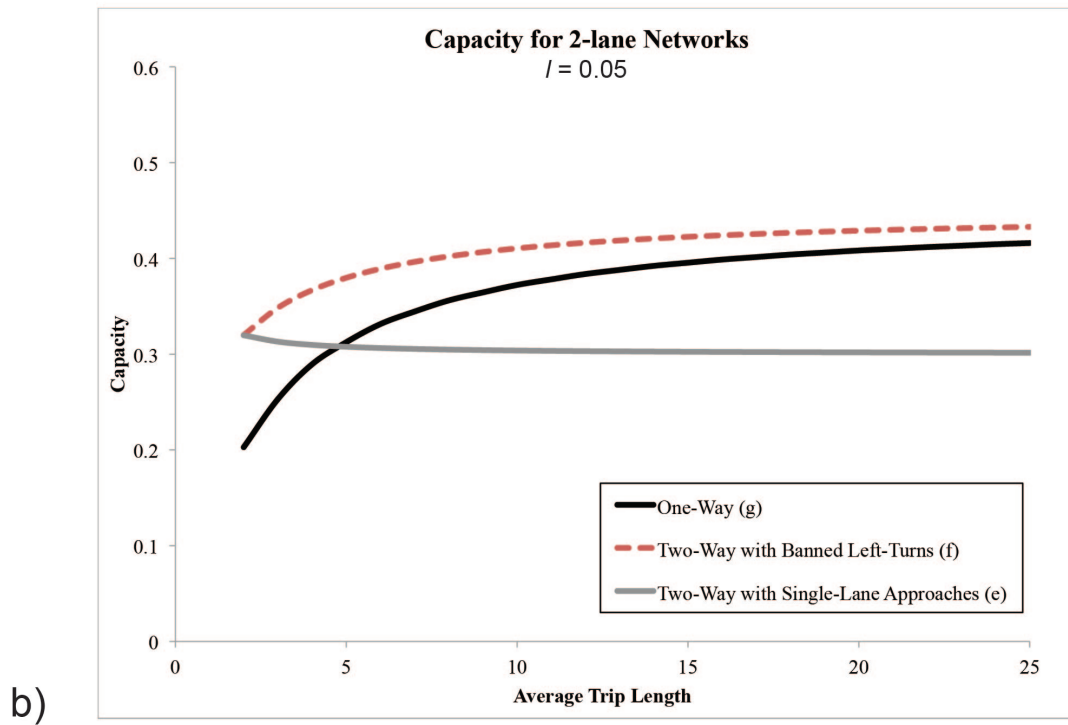
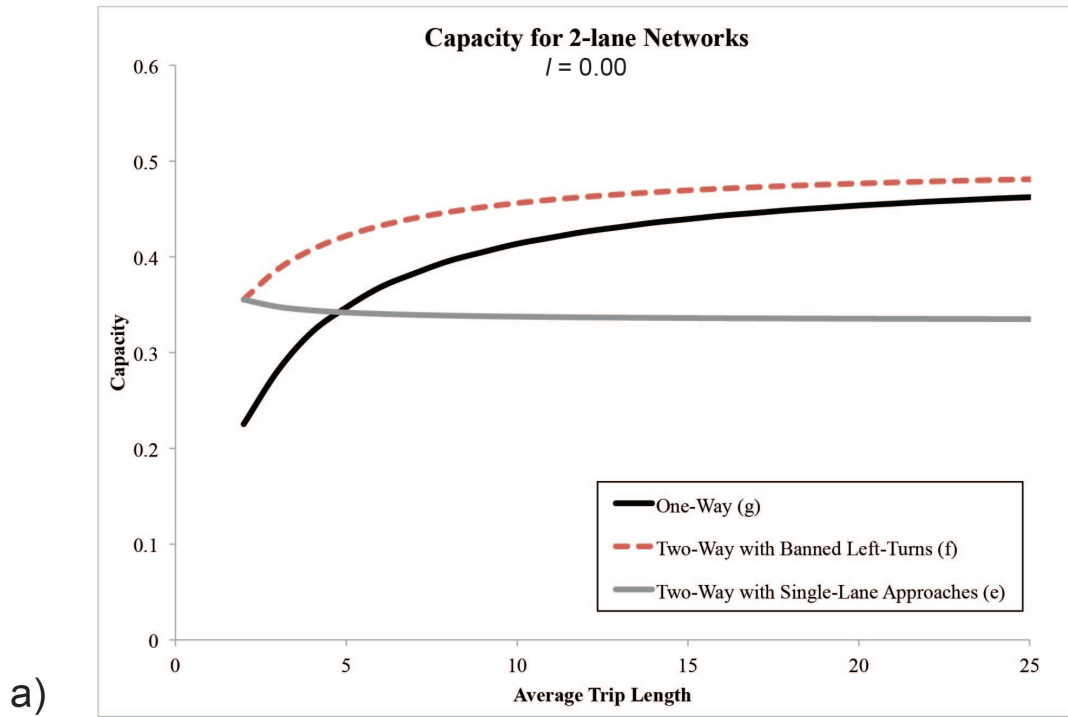


Figure 5.3. Capacities of 2-lane networks for: (a) $l = 0.00$; and, (b) $l = 0.05$.

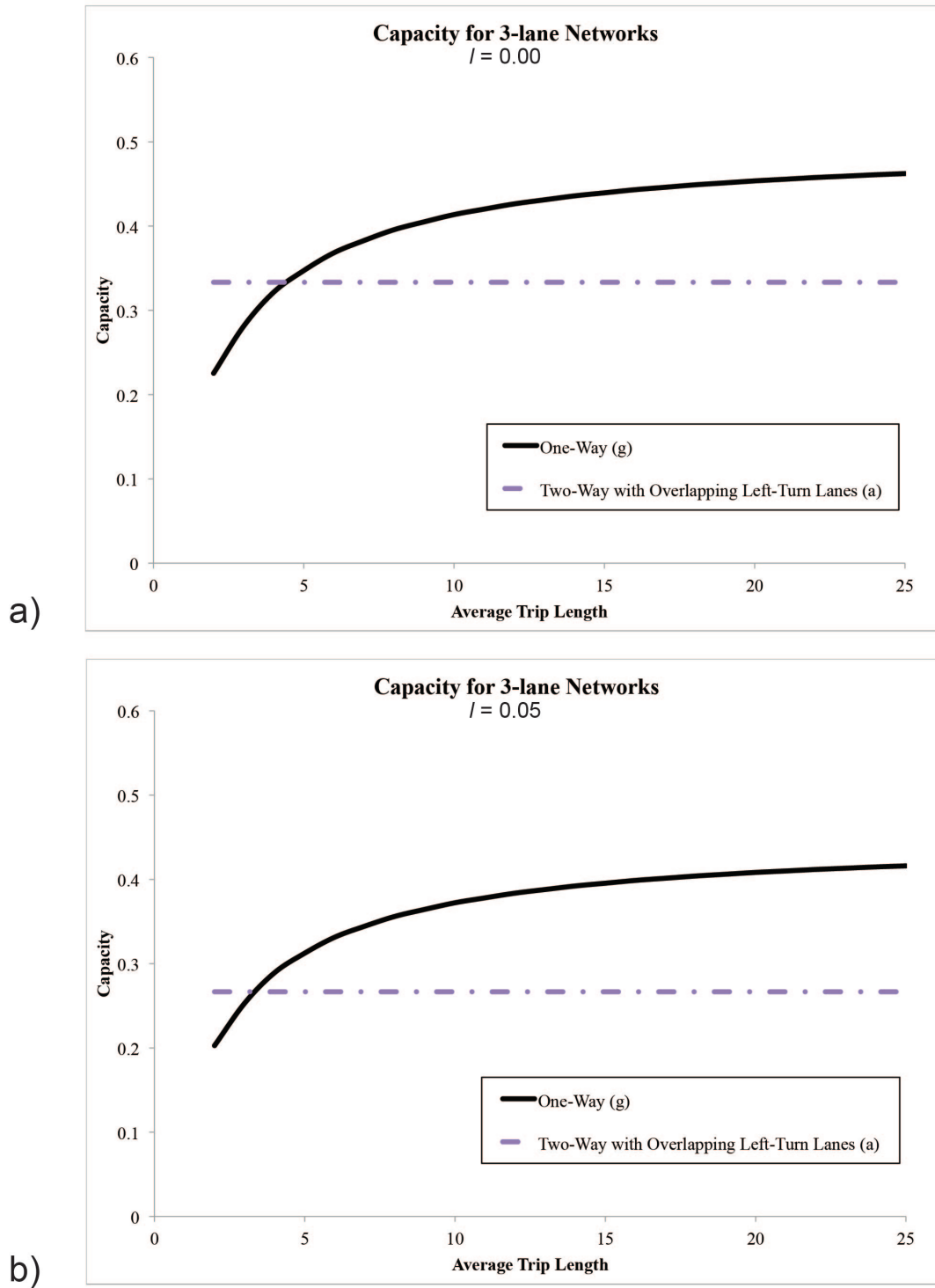


Figure 5.4. Capacities of three-lane networks for: (a) $l = 0.00$; and, (b) $l = 0.05$.

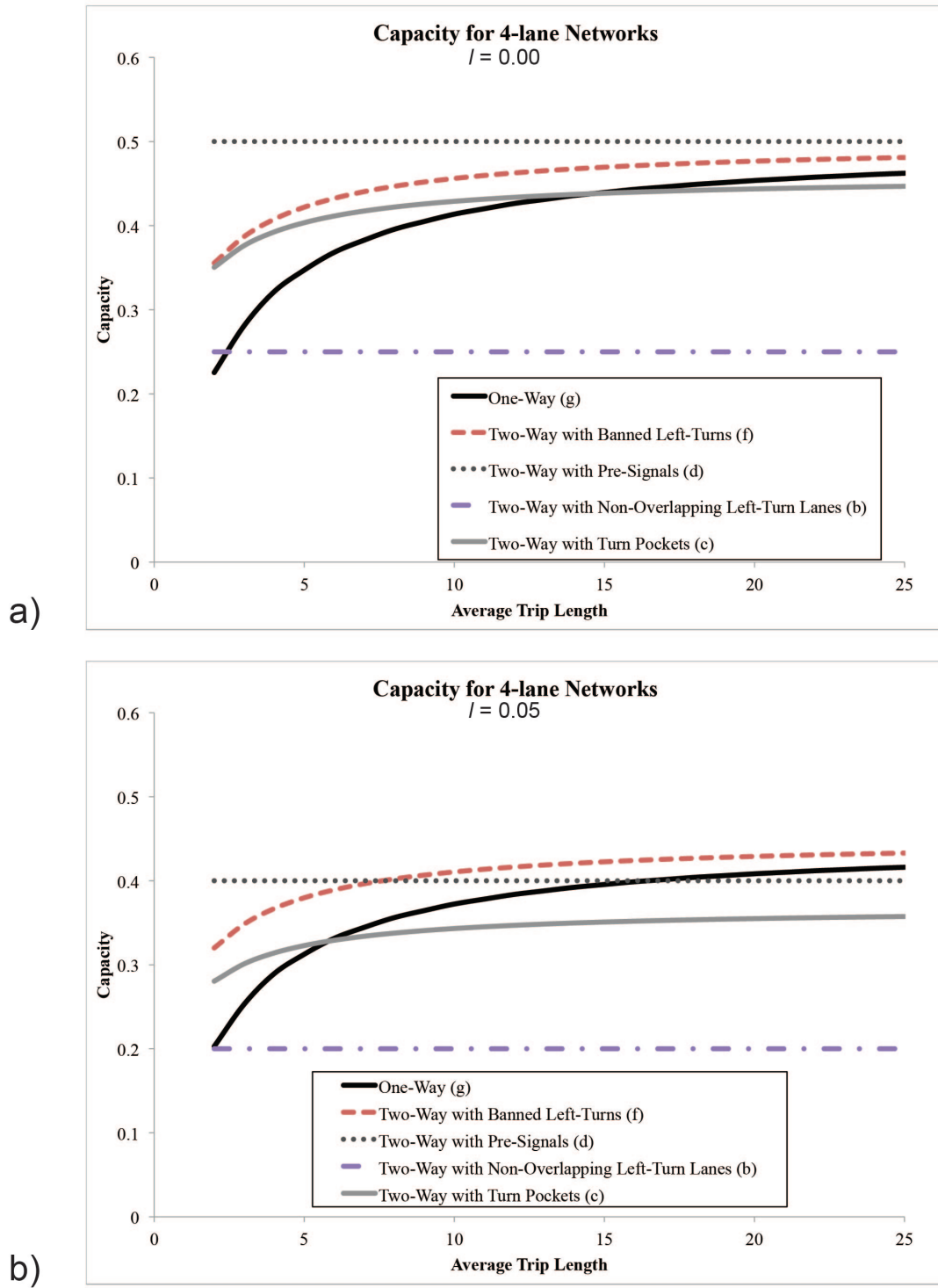


Figure 5.5. Capacities of four-lane networks for: (a) $l = 0.00$; and, (b) $l = 0.05$.

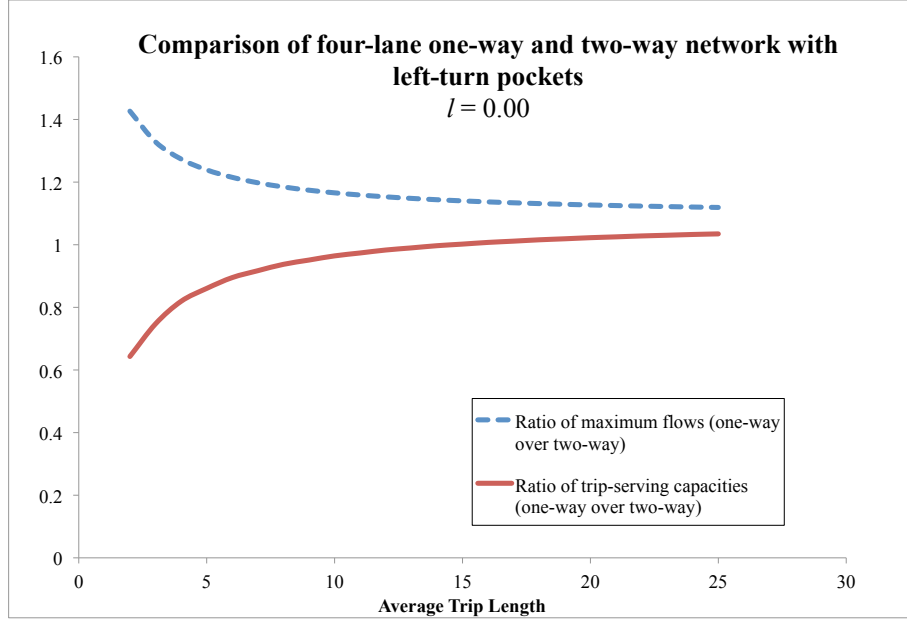


Figure 5.6. Comparison of four-lane one-way network and two-way network with left-turn pockets.

that uses the same space but provides turn pockets to serve conflicting left turns. Notice that maximum vehicle flows are always higher on the one-way network, but the two-way network with left-turn pockets has a higher trip-serving capacity when trips are short ($n < 14$). Also, trip-serving capacities of the two-way network are never less than 3.5% of that provided by the one-way network for the entire range of trip lengths presented.

This trend exists for all the networks examined. For networks consisting of two travel lanes, the two-way network with single lane approaches can serve more trips per unit time than the one-way network when average trips lengths are less than 4 blocks. For three-lane networks, the two-way network with overlapping left-turn lanes also outperforms the one-way network for short trips ($n < 4$ for $l = 0.00$ and $n < 3$ for $l < 0.05$). For four-lane networks, the two-way network with left-turn pockets is again superior to the one-way network for short trips, especially when lost time is negligible ($n < 14$ for $l = 0.00$ and $n < 6$ for $l = 0.05$). Since most downtown areas are small and trips can be expected to be short, it is therefore possible that allowing conflicting turning maneuvers on two-way streets is better for overall network performance than using networks of one-way streets.

Non-traditional treatments of conflicting left turns can also provide superior network performance when compared to using one-way streets. The two-way network with pre-signals outperforms the one-way network for all trip lengths when $l = 0.00$ and for shorter trips when $l = 0.05$ ($n < 17$). Additionally, the two-way network with banned left-turns always outperforms the one-way network, even if trips are long. This is because both the one-way network and two-way network with banned

left turns provide the same saturation flows and amount of lane-seconds of green for vehicles to discharge (i.e., same vehicle-moving capacity), but the latter imposes less circuitous routes than the former. Therefore, if trips are long the most beneficial treatment for network performance is to use two-way streets and simply ban left turns at intersections. This treatment provides large maximum vehicle flows, relatively short average trip lengths and does not require costly infrastructure investments like left-turn pockets, pre-signals or one-way street configurations.

The results presented assume that signal settings are the same for networks that use a two-phase and four-phase signal. However, signals with more phases tend to have longer cycle lengths to reduce the fraction of discharge time lost due to phase changes. While not done here, the model can easily be used to compare networks with different cycle lengths by varying the value of l between these networks. Based on the results presented here, a smaller value of l will make networks that allow conflicting left turns in protected phases (and have more complicated phasing schemes) have capacities that are more competitive with those that do not.

5.5 Extension to more realistic networks

The previous analysis relies on some simplifying assumptions that may not hold in reality. This section discusses how the results of this analysis methodology might change when some of these assumptions are relaxed. This includes the assumptions that: lost times and saturation flows are the same for left-turn through movements, left-turn phases are perfectly timed to serve the left-turn demand exactly and that traffic is evenly balanced between the north-south, and east-west directions.

This analysis assumes equal lost times for all phase changes. In reality, the lost time incurred after a phase depends on the movement served by that phase. For example, Bonneson (1992) shows that the lost times incurred after left-turning movements are greater than those incurred after through movements. This can easily be incorporated into the model by changing the fractions of the cycle available to serve each street family in (5.2), (5.6) and (5.7).

It was also assumed here that through and left-turning vehicles discharge through the intersection at the same rate. However, the literature suggests that the saturation flows of left-turning vehicles in exclusive lanes is at most 0.95 times that of through vehicles (HCM, 2000). This can also be easily incorporated into the capacity calculations by modifying (5.1) so that a different saturation flow ratio is used for left-turn and through-moving vehicles. In this case, (5.1) changes to:

$$C_j = \left[\frac{N_j^T g_j^T s_j^T + N_j^L g_j^L s_j^L}{N_j} \right] \frac{1}{\alpha_j}, \quad (5.14)$$

where s_j^T and s_j^L represent the saturation flow ratio for through-moving and left-turning vehicles, respectively. Both of these changes will reduce the capacity of the networks that allow conflicting left-turning maneuvers. While this would not change

the general conclusion that allowing these maneuvers might be beneficial when trips are short, it would reduce the critical trip length at which banning left turns would provide better network performance.

This analysis also assumes that the left-turn phases are perfectly timed to serve the average left-turning demand. In reality, however, traffic engineers use a minimum left-turn phase length (usually about 7 to 10 seconds long) to accommodate fluctuation in this demand at one location between cycles. Thus, the capacity here is overestimated for situations in which the perfectly timed left-turn phase length is shorter than this minimum length. However, minimum left-turn phase lengths can easily be incorporated into the capacity analysis. If the minimum green fraction for the left-turn phase is denoted g_j^{L*} , (5.14) can be rewritten as:

$$C_j = \begin{cases} [N_j^T g_j^T s_j^T + N_j^L g_j^L s_j^L] / [N_j \alpha_j] & \text{if } g_j^L \geq g_j^{L*} \\ [N_j^T g_j^T s_j^T + z N_j^L g_j^{L*} s_j^L] / [N_j \alpha_j] & \text{otherwise} \end{cases}, \quad (5.15)$$

where z is the average proportion of the minimum left-turn phase that is actually used by vehicles to discharge. When minimum green times for the left-turn phase are included, the capacity of two-way networks will only be affected (and reduced) when g_i^L is small. Since (5.4) shows that g_i^L decreases with n , this will only occur when n is large, which happen to be cases where eliminating conflicting left-turns is already advantageous. Therefore, the previous insights that allowing conflicting left-turns on two-way street networks is beneficial when average trip lengths are short should still hold.

It was also assumed that origins and destinations were evenly distributed so that travel was evenly split between the two families of streets. However, in some networks the distribution of origins and destinations may be such that demand is biased in favor of one street family. For example, vehicles may travel more in the north-south direction than in the east-west direction, on average. In this case, it would not make sense to evenly split the green time between the two street families. Instead, the direction with more demand should receive more green time. The same type of analysis shows that if the green time is proportioned correctly to account for the asymmetric demand, the capacities given by (5.6) and (5.7) and presented in Figures 5.3, 5.4 and 5.5 will not change if demand is unbalanced. Thus, the previous results should still hold if signal timings are optimally adjusted to accommodate different demand patterns.

5.6 Major findings and discussion

The model presented here can be used to quantitatively compare the trip-serving (and vehicle-moving) capacities of homogeneous grid networks based on the treatment of conflicting left turns. These capacities turn out to be a function of just a few (measurable) characteristics: the number, type and width of the lanes available for discharge at intersections, the fraction of the signal cycle lost for vehicle movement

between phase changes and the average trip length. This model can be used as a tool for comparing different networks, or when planning conversions across a particular network (e.g., a one-way to two-way street conversion network wide).

The model unveils several general insights into how turns affect the operation of urban traffic networks. Perhaps the most surprising insight contradicts conventional wisdom and design handbooks that suggest one-way networks are always more efficient than two-way networks because the former offer higher vehicle flows (Pline, 1992; HCM, 2000). While one-way streets offer higher vehicle flows across the network because they eliminate conflicting left turns, comparing the more pertinent metric of trip-serving capacities shows that one-way networks serve trips at a lower rate than comparable two-way networks when trips are short. This is because the additional circuitry imposed by one-way networks more than offsets the higher vehicle flows obtained by eliminating left turns. Two-way networks that allow conflicting left-turns also become more competitive with one-way networks as the fraction of the signal cycle lost per phase change decreases (or as the length of the cycle increases). Since two-way networks that allow conflicting left-turns typically tend to have longer cycles in reality, they should be even more competitive with one-way networks than suggested in the results presented here.

This work also shows that two-way networks would always provide higher trip-serving capacities than comparable one-way networks if left-turns are banned, even if trips are long. While both strategies eliminate conflicting left-turn maneuvers so that intersections can be served with two-phase traffic signals (and, therefore, provide the same vehicle-moving capacity), two-way networks with banned left-turns impose less circuitry than one-way networks. Thus, a one-way network can always be converted to two-way operation while increasing the rate that trips can be served if left turns are simply banned at intersections.

This finding is especially important given that many cities have recently considered converting traditional one-way downtown streets to two-way operation (Dorroh & Kochevar, 1996; Hart, 1998; Schwinger *et al.*, 1999; Lyles *et al.*, 2000; Hawkins *et al.*, 2006; Vo *et al.*, 2007). Two-way streets are seen as desirable for a variety of reasons (Walker *et al.*, 2000): they are less confusing than one-way streets; more conducive to economic activity and a livable environment (Forbes, 1998); potentially safer (Tindale & Hsu, 2005); and, they allow vehicles (especially transit and emergency vehicles) to take more direct routes. However, the main hesitation in making these conversions is a perceived loss in network efficiency. The model presented here, however, shows that these concerns may not be warranted. If trips are short, two-way streets that allow conflicting turns can still outperform comparable one-way networks during a rush hour period. And if trips are long, conflicting left-turns can simply be banned at the intersection to provide superior efficiency. Thus, this work shows that livability and efficiency objectives can be achieved simultaneously.

Chapter 6

Conclusions

This chapter provides concluding remarks. Section 6.1 provides a summary of the major findings of this work. Section 6.2 discusses ways in which this research can be extended in the future.

6.1 Summary of major findings

This dissertation contributes to the current body of knowledge on the aggregate (i.e., system-wide) behavior of vehicles on urban traffic networks by examining the effects of the turning maneuvers that exist in networks with multiple overlapping routes. This work shows that the presence of multiple routes and turning maneuvers affects both the stability of traffic on the network and the trip-serving capacity of the network itself.

The first part of this dissertation examined the destabilizing effects of turning maneuvers on multi-route networks. It showed that when average network densities are low, traffic naturally tends toward equilibrium states in which vehicles are uniformly distributed across all links. This yields a consistent and robust relationship between average network flow and average network density, equivalent to the MFD that is expected from current macroscopic theory. However, as average densities increase past a critical point, uniform vehicle distributions no longer remain sustainable. Traffic will then naturally tend toward equilibrium states in which vehicles are unevenly distributed across the network. This uneven distribution causes unpredictable aggregate behavior and in some cases can lead to complete gridlock, well below maximum (jam) density. Fortunately, the presence of drivers that adaptively navigate away from localized pockets of congestion in real-time mitigates this destabilizing effect. This highlights the importance of designing networks that are redundant (i.e., have multiple routes that serve every origin-destination pair) and providing drivers with real-time information to avoid localized congestion. Since this instability only arises when the network becomes congested, this work also highlights the need to use macroscopic control strategies (e.g., pricing and metering schemes) to keep densities from exceeding a critical value.

This dissertation also showed how this natural instability affects network dynamics. The work unveiled that vehicles on multi-route urban networks naturally tend to be more uniformly distributed across all links as average densities increase (i.e., the onset of congestion) than as average densities decrease (i.e., the dissipation of congestion). This imbalance occurs because vehicles exiting the network clear more slowly from more congested regions. Since uneven vehicle distributions reduce average network flows, it follows that lower network flows will naturally arise as average network densities decrease than as average density increases and clockwise loops would appear in plots of average flow vs. average density. Again, the presence of adaptive drivers helps to mitigate this effect. The result is that these loops should only be expected when network-wide disturbances hinder the ability of drivers to adapt, as verified with empirical data.

The second part of this dissertation examined how the presence of turning maneuvers can reduce a network’s trip-serving capacity. The analysis focused on conflicting turning maneuvers at intersections and found that efforts to mitigate these conflicts on two-way streets (e.g., separate turning lanes or phases at intersection signals) reduce the vehicle-moving capacity of the network. Another mitigation strategy is to design the network as a system of one-way streets, which eliminates conflicting turning maneuvers entirely and allows the highest vehicle flows. However, an increased ability to serve vehicle flows does not necessarily equate to an increased ability to serve vehicle-trips; the latter is a more pertinent metric to describe network performance during a rush hour period. This analysis used macroscopic analysis techniques to compare the trip-serving capacities of different network types based on the treatment of conflicting left turns. It recognized that some treatments require movement or turning restrictions that may force vehicles to travel longer distances on average, which decreases the trip completion rate. This work found that, contrary to the existing literature and design handbooks, two-way networks that allow conflicting left-turn maneuvers can serve trips at a higher maximum rate than comparable one-way networks when average trip lengths are short. Additionally, two-way networks that ban conflicting left-turns are able to serve trips at a higher maximum rate than one-way networks, even when trips are long. Since urban planners prefer two-way networks because they are more conducive to thriving downtown environments, these results show that livability and efficiency objectives can be achieved simultaneously in many cases.

6.2 Future work

There are several ways the work presented in this dissertation can be extended to improve our understanding of how traffic behaves on urban networks. Some of these future research areas are:

1. Determining the most appropriate ways to improve driver adaptivity on a network. This research has found that the presence of drivers that are willing and able to make routing decisions in real time to avoid congested areas helps im-

prove aggregate network behavior. However, the research does not necessarily provide insights into how driver adaptivity can be improved in the field. Future work may consider how to use fixed and mobile probes to determine conditions within the network in real time, and develop routing algorithms that can help drivers avoid congested areas. Additional work can be done to determine the best way to relay this information to drivers, particularly those who are already in-route. Interacting with in-vehicle navigation systems and ITS technologies such as variable message signs and speed limits seem like promising avenues to explore.

2. Comparing the efficiency of undersaturated networks based on the treatment of conflicting left turns. This dissertation examines the effect of different conflicting left turn treatments on the trip-serving capacity of a network by examining network operation at saturation. However, most networks operate at saturation for a small portion of the day, and some do not have high enough demands to reach saturation at all. It is possible that treatments aimed at improving efficiency at saturation will decrease efficiency when the network is under-saturated. Future work may consider the effect of different conflicting left turn treatments when the network is operating in light traffic conditions, and how these treatments can be changed in real time as demand increases and the network approaches saturation.
3. Investigating the effects of network hierarchy on the aggregate behavior of vehicles traveling in urban areas. Many urban networks have a hierarchical structure in which a few high-capacity arterials streets are served by many low-capacity collectors. In this dissertation, this hierarchical structure is unaccounted for because streets were generally assumed to be completely homogeneous. In the few cases where network heterogeneity was discussed, this dissertation only considered some simple cases where all roadways serving one travel direction were identical. Future work may consider how hierarchical structures affect aggregate traffic relationships, and use this information to determine the most efficient hierarchical network structures.
4. Exploring the behavior of multimodal urban networks. This dissertation focuses on the behavior of a single mode, cars, on urban streets. However, most networks also contain a variety of other modes that interact on urban streets: buses, passenger rail, bicycles, pedestrians and urban freight vehicles. Future work may consider how the presence of these different modes can affect aggregate network behavior. This may start at a local level (e.g., a single intersection or link) to develop insights on the types of interactions and conflicts that emerge. Once a thorough understanding is obtained at a local level, this information might then be scaled up to describe the network-wide behavior of multimodal traffic.

Bibliography

2000. *Highway Capacity Manual*. Transportation Research Board.
- Ardekani, S., & Herman, R. 1987. Urban network-wide traffic variables and their relations. *Transportation Science*, **21**(1), 1–16.
- Bonneson, J. 1992. Study of headway and lost time at single-point urban interchanges. *Transportation Research Record*, 30–39.
- Buisson, C., & Ladier, C. 2009. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *Transportation Research Record*, **2124**, 127–136.
- Cervero, R. 2002. Induced travel demand: Research design, empirical evidence, and normative policies. *Journal of Planning Literature*, **17**(1), 3–20.
- Cottrell, B.H. 1986. Guidelines protected/permissive left-turn signal phasing. *Transportation Research Record*, **1069**, 54–61.
- Daganzo, C.F.. 1998. Queue spillovers in transportation networks with a route choice. *Transportation Science*, **32**(1), 3–11.
- Daganzo, C.F. 2006. In traffic flow, Cellular Automata = Kinematic Waves. *Transportation Research Part B*, **40**(5), 396–403.
- Daganzo, C.F. 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B*, **41**(1), 49–62.
- Daganzo, C.F., & Geroliminis, N. 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B*, **42**(9), 771–781.
- Daganzo, C.F., Gayah, V.V., & Gonzales, E.J. 2011. Macroscopic Relations of Urban Traffic Variables: Bifurcations, multivaluedness and instability. *Transportation Research Part B*, **45**(1), 278–288.
- Daganzo, C.F., Gayah, V.V., & Gonzales, E.J. 2012. The potential of parsimonious models for understanding large scale transportation systems and answering big picture questions. *EURO Journal of Transportation and Logistics*, doi 10.1007/s13676-012-0003-z.

- Dorroh, R., & Kochevar, R. 1996. One-way conversion for calming Denver's streets. *Pages 109–113 of: Proceedings of the 1996 ITE International Conference.*
- Eddie, L.C. 1965. Discussion of traffic stream measurements and definitions. *Pages 139–154 of: Proceedings of the 2nd International Symposium on the Theory of Traffic Flow.*
- Eliasson, J., & Mattsson, L.-G. 2006. Equity effects of congestion pricing and a case study for Stockholm. *Transportation Research Part A*, **40**(7), 602–620.
- Fambro, D.B., Messer, C.J., & Anderson, D.A. 1977. Estimation of unprotected left-turn capacity at signalized intersections. *Transportation Research Record*, **644**, 113–119.
- Forbes, G. 1998. Vital signs: Circulations in the heart of the city—an overview of downtown traffic. *ITE Journal*, **68**(8), 26–29.
- Geroliminis, N., & Daganzo, C.F. 2007. Macroscopic modeling of traffic in cities. *In: Proceedings of the 86th Annual Meeting of the Transportation Research Board.*
- Geroliminis, N., & Daganzo, C.F. 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B*, **42**(9), 759–770.
- Geroliminis, N., & Sun, J. 2011a. Hysteresis phenomena of a Macroscopic Fundamental Diagram in freeway networks. *Transportation Research Part A*, **45**(9), 966–979.
- Geroliminis, N., & Sun, J. 2011b. Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B*, **45**(3), 605–617.
- Godfrey, J. 1969. The mechanism of a road network. *Traffic Engineering and Control*, **11**(7), 323–327.
- Gonzales, E.J., Chavis, C., Li, Y., & Daganzo, C.F. 2009. Multimodal transport modeling for Nairobi, Kenya: Insights and recommendations with an evidence based model. *UC Berkeley, Volvo Working Paper UCB-ITS-VWP-2009-5.*
- Hart, J. 1998. Converting back to two-way streets in downtown Lubbock. *ITE Journal*, **68**(8), 36, 45–46.
- Hawkins, N., Isebrands, H., & Fox, G. 2006. Converting CBD roadways from one-way to two-way traffic flow: A case study for the City of Des Moines, Iowa. *In: 2006 ITE Annual Meeting and Exhibit Compendium of Technical Papers.*
- Herman, R., & Prigogine, I. 1979. A two-fluid approach to town traffic. *Science*, **204**(4389), 148–151.

- Jin, W.-L. 2009. Asymptotic traffic dynamics arising in diverge-merge networks with two intermediate links. *Transportation Research Part B*, **43**(5), 575–595.
- Lee, D., Klein, L., & Camus, G. 1999. Induced traffic and induced demand. *Transportation Research Record*, **1659**, 68–75.
- Lighthill, M.J., & Whitham, G.B. 1955. On kinematic waves. I: Flow movement in long rivers. II: A Theory of traffic flow on long crowded roads. *Proceedings of the Royal Society*, **229**(A), 281–345.
- Litman, T. 2011. *London congestion pricing: Implications for other cities*. Tech. rept. Victoria Transport Policy Institute.
- Lyles, R., Faulkner, C., & Syed, A. 2000. Conversion of streets from one-way to two-way operation. *Final report for the Michigan Department of Transportation*.
- Mahmassani, H., & Peeta, S. 1993. Network performance under system optimal and user equilibrium dynamic assignments: Implications for advanced information systems. *Transportation Research Record*, **1408**, 83–93.
- Mahmassani, H., Williams, J., & Herman, R. 1987. Performance of urban traffic networks. *Pages 1–20 of: Proceedings of the 10th International Symposium on Transportation and Traffic Theory*.
- Mazlounian, A., Geroliminis, N., & Helbing, D. 2010. The spatial variability of vehicle densities as determinant of urban network capacity. *Philosophical Transactions of the Royal Society*, **368**(1928), 4627–4647.
- Messer, C.J., & Fambro, D.B. 1977. Effects of signal phasing and length of left-turn bay on capacity. *Transportation Research Record*, **644**, 95–101.
- Nagel, K., & Schreckenberg, M. 1992. A cellular automaton model for freeway traffic. *Journal de Physique I*, **2**(12), 2221–2229.
- Nash, A.B., & Sylvia, R. 2001. *Implementation of Zurich’s transit priority program*. Tech. rept. Mineta Transportation Institute.
- Newell, G.F. 1959. The effect of left turns on the capacity of a traffic intersection. *Quarterly of Applied Mathematics*, **17**(1), 67–76.
- Olszewski, P., Fan, H., & Tan, Y. 1995. Area-wide traffic speed-flow model for the Singapore CBD. *Transportation Research Part A*, **29**(4), 273–281.
- Phang, S.Y., & Toh, R. 2004. Road congestion pricing in Singapore: 1975 to 2003. *Transportation Journal*, **43**(2), 16–25.
- Pline, J. (ed). 1992. *Traffic Engineering Handbook*. 4th edn. Prentice Hall.

- Richards, P.I. 1956. Shock waves on the highway. *Operations research*, **4**(1), 42–51.
- Schrank, D., & Lomax, T. 2009. *2009 Urban Mobility Report*. Tech. rept. Texas Transportation Institute.
- Schwinger, C., O’Leary, S., Fulton, T., & Graham, G. 1999. Central business district traffic circulation improvement—a case study, Salina, KA. *In: Proceedings of the 68th Annual Meeting of the Institute of Transportation Engineers*.
- Smeed, R. 1966. Road capacity of city centers. *Traffic Engineering and Control*, **9**(7), 455–458.
- Smeed, R., & Wardrop, J. 1964. An exploratory comparison of the advantages of cars and buses for travel in urban areas. *Journal of the Institute of Transport*, **30**(9), 301–305.
- Thomson, J.M. 1967. Speeds and flows of traffic in central London: 2. Speed-flow relations. *Traffic Engineering and Control*, **8**(12), 721–725.
- Tindale, S., & Hsu, P. 2005. Crash data and signal coordination: A one-way pair case study. *Journal of Safety Research*, **36**(5), 481–482.
- Vo, P., Mattingly, S., Ardekani, S., & Dilshad, Y. 2007. Comparison of quality of service in two central business districts: Two-fluid model approach in Texas. *Transportation Research Record*, **1999**, 180–188.
- Walker, G., Kulash, W., & McHugh, B. 2000. Downtown streets: Are we strangling ourselves on one-way networks? *Transportation Research Circular*, **501**.
- Wardrop, J. 1968. Journey speed and flow in central urban areas. *Traffic Engineering and Control*, **9**(11), 528–532, 539.
- Wong, C.K., & Wong, S.C. 2003. Lane-based optimization of signal timings for isolated junctions. *Transportation Research Part B*, **37**(1), 63–84.
- Xuan, Y., Gayah, V.V., Daganzo, C.F., & Cassidy, M.J. 2010. Multimodal Traffic at Isolated Signalized Intersections: New Management Strategies and a Framework for Analysis. *In: Proceedings of the 89th Annual Meeting of the Transportation Research Board*.
- Xuan, Y., Daganzo, C.F., & Cassidy, M.J. 2011. Increasing the capacity of signalized intersections with separate left turn phases. *Transportation Research Part B*, **45**(5), 768–781.
- Xuan, Y., Gayah, V.V., Daganzo, C.F., & Cassidy, M.J. 2012. Using a pre-signal to increase intersection capacity: Theory and experiment. *Transportation Research Record*. [in press].

Zahavi, Y. 1972. Traffic performance evaluation of road networks by the α -relationships. Parts 1 and 2. *Traffic Engineering and Control*, **14**(5 and 6), 228–231, 292–293.