Modeling and Software Tools for Freeway Operational Planning

Alex A. Kurzhanskiy

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2007-148
http://www.eecs.berkeley.edu/Pubs/TechRpts/2007/EECS-2007-148.html

December 13, 2007

Acknowledgement

I would like to thank TOPl group members, Pravin Varaiya, Roberto Horowitz,
Jaimyoung Kwon and Gabriel Gomes for making this work possible;
Ram Rajagopal for sharing his ideas regarding development of Aurora framework;
and Saurabh Amin for fruitful discussions about conservation laws.

Modeling and Software Tools for Freeway Operational Planning

by

Alexandr A. Kurzhanskiy

Diploma (Lomonosov Moscow State University, Russia) 2000

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering and Computer Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Pravin Varaiya, Chair
Professor Roberto Horowitz
Professor Alexander Skabardonis
Professor Claire Tomlin

Fall 2007

The dissertation of Alexandr A. Kurzhanskiy is approved:

Chair Date

Date

Date

Date

University of California, Berkeley

Fall 2007

Modeling and Software Tools for Freeway Operational Planning

by

Alexandr A. Kurzhanskiy

Abstract

Modeling and Software Tools for Freeway Operational Planning

by

Alexandr A. Kurzhanskiy

Doctor of Philosophy in Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Pravin Varaiya, Chair

This dissertation grew out of the author's participation in the *Tools for Operational Planning* (TOPl) project. TOPl project started at Berkeley in April 2006. Its purpose is to provide tools for quantative assessment of operational strategies designed to improve traffic conditions on congested freeways and surrounding arterials. The elements of such strategies are:

demand management, which focuses on reducing excessive demand;

incident management, which targets resources to alleviate accident hot spots;

traveler information, which potentially reduces traveler buffer time; and

traffic control, which implements aggressive ramp metering at locations where significant

reductions in congestion are likely to occur.

TOPl models traffic flows on road networks in realistic conditions relying on freeway data provided by the California freeway *Performance Measurement System* (PeMS). PeMS offers an accounting framework for tracking freeway performance, and a suite of diagnostic tools that pinpoint the weaknesses in freeway operations. For arterial data, TOPl uses various alternative sources, such as regional planning agencies—MTC (Bay Area) and SANDAG (San Diego) for GIS description of road networks, and census information and demographics surveys for origin-destination travel patterns and demand.

Macroscopic traffic models represent traffic as a compressible fluid in terms of flow, density and speed, as opposed to *microscopic* models which seek to reproduce the behavior of individual vehicle as its driver responds to its environment by changing its speed and lane. TOPl uses the macroscopic approach, as it is based on sound theory, is easy to implement in software, and the implementations are fast to run, allowing the user to simulate many different traffic situations in relatively short time. Our model of choice is the *cell transmission model* (CTM)—a special case of Godunov discretization of the Lighthill-Whitham-Richards first order model, with triangular fundamental diagram. While simple, CTM adequately describes traffic flow on freeways, and the simulation results match well the measurement data provided by PeMS. Exploring the CTM model from the point of view of nonlinear dynamical systems we describe the structure of its equilibrium points and the behavior of its trajectories under different demand patterns. Then we discuss the implications of our findings for ramp metering.

CTMSIM is the interactive MATLAB based freeway traffic simulator of the CTM model.

It allows plugging in user-defined ramp flow and ramp queue controllers, its output results can be directly compared with PeMS data, and it can operate in both graphical (interactive) mode and command line (batch) mode. Simple and robust, CTMSIM has proved to be a handy tool for transportation researchers who can use it for evaluating ramp metering algorithms and for estimating the impact of different response times in the incident management. CTMSIM can only deal with a single freeway. It does not support arterials, freeway networks, or even HOV lanes.

The *Aurora* object-oriented framework overcomes the limitations of the CTMSIM. Its basic objects, nodes and links, allow the user to construct heterogeneous road networks. Various event classes make it possible to generate simulation scenarios. The monitor objects can keep track of the state at selected nodes and links, coordinate control actions at nodes, or generate events at nodes or links when the monitored states reach certain thesholds. Monitors and events enable the modeling of the impact of traveler information as well as incident management and the coordination of signal control on arterials with the ramp metering at freeway entries. The analysis module of Aurora, which is still under development, will address the issue of demand management: the goal is to solve the user equilibrium dynamic traffic assignment problem and evaluate various toll mechanisms.

There are other benefits of Aurora.

It is scalable: the nodes of Aurora network can be networks themselves allowing the user to build the network configurations incrementally.

It is designed to be multi-purpose: the basic infrastructure and algorithms are generic and not specific to transportation; link dynamics is implemented as interface allowing to plug

in different models (currently only CTM is implemented). Thus, it can be used not just for modeling traffic on road networks but also for other applications, such as irrigation canals, oil or gas pipelines, etc.

Finally, the Aurora simulator is a standalone Java application, and as such it does not require MATLAB.

Professor Pravin Varaiya
Dissertation Committee Chair

To my mama and papa.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank TOPl group members, Pravin Varaiya, Roberto Horowitz, Jaimyoung Kwon and Gabriel Gomes for making this work possible; Ram Rajagopal for sharing his ideas regarding development of Aurora framework; and Saurabh Amin for fruitful discussions about conservation laws.

Chapter 1

Introduction

The Highway Safety, Traffic Reduction, Air Quality, and Port Security Bond Act of 2006, approved by the voters of California as Proposition 1B on November 7, 2006, includes a program of funding from $4.5 billion to be deposited in the Corridor Mobility Improvement Account (CMIA). The basic objective of CMIA is to improve performance on highly congested travel corridors[1].

Improvements that relieve congestion by expanding capacity, enhancing operations, or otherwise improving travel times or reducing the number of daily vehicle-hours of delay within high-congestion travel corridors, may be on the state freeway system or on major access routes to the state freeway system.

These improvements are addressed by the Transportation Management System (TMS) Master Plan [99] whose system management philosophy is summarized in Figure 1.1. The pyra-

[1]Road networks comprised of freeways and urban arterials

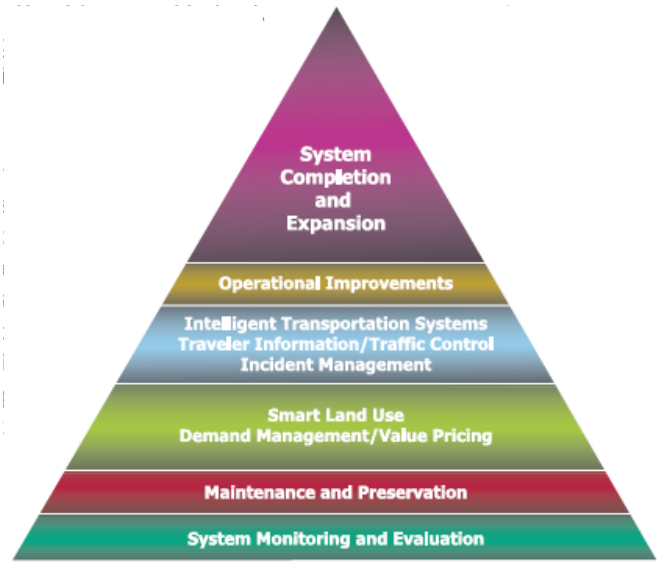


Figure 1.1: System management philosophy. Source: [99]

mid rests on the foundation of "System Monitoring and Evaluation", which in considerable part is provided by the freeway *Performance Measurement System* (PeMS) [1][2].

Above this base is the "Maintenance and Preservation" layer, emphasizing timely intervention to economically maintain the capital assets of the State's freeway system. The middle layers of the pyramid concern freeway operations strategies—demand management, incident management, traveler information and traffic control—that offer large productivity gains at low cost. The two layers at the top concern capital expenditures: small investments in operational improvements, such as increased ramp capacity or an extended auxiliary lane;

[2]PeMS provides an accounting framework for tracking California freeway performance, and a suite of diagnostic tools that pinpoint the weaknesses in freeway operations. PeMS was initially to be a simple system to calculate the most basic performance measures using available sensor measurements and processing algorithms; incur low cost; and permit incremental implementation [106]. Over several years PeMS has evolved into a unique database of freeway data, going back to 1998, together with a suite of tools that help analyze these data at different scales of aggregation and for variety of performance measures. At the scale of freeway, district or the entire state, PeMS estimates freeway productivity and congestion; travel time reliability; and the location, severity and frequency of bottlenecks and collision "hot spots". At the micro-level, PeMS applications reveal the impact of an individual collision, lane closure or special event.

and major investments in long-term system expansion.

The *Tools for Operational Planning* (TOPl) project started at Berkeley in April 2006[3].

Its research is guided by the TMS Master Plan, so its analysis tools match the Action Plans that the Master Plan envisages for the middle layers of the pyramid. Accordingly,

- the "demand management" layer focuses on reducing "excess demand";
- the "incident management" layer targets resources to alleviate accident hot spots;
- the "traveler information" layer seeks to reduce traveler buffer time; and
- the "traffic control" layer implements aggressive ramp metering at locations where the maximum reductions in congestion are likely to occur.

The objective of TOPl is to provide a *quick quantative assessment* of the benefits that can be gained from these TMS Action Plans.

The quick quantative assessment provided by TOPl, can help *rank* a large set of Action Plans in terms of the benefits they will yield. Combined with a separate estimate of the cost of these plans, TOPl can serve as the first step in selecting the most promising of them. This initial selection may be based on benefit/cost ratios or the magnitude of benefits.

The focus of TOPl is on operations in freeway corridors (road network comprised of freeway and surrounding arterials). A corridor is the smallest spatial unit that can be consistently

[3]TOPl is supported by the California Department of Transportation through the California PATH program.

analyzed as a self-contained system. Suppose, for example, that we wish to consider the impact of a promising new metering algorithm on some ramps on a given freeway. Evidently, this impact will depend on how other ramps on this freeway are metered. Furthermore, the impact of metering will affect (and be affected by) the signaling strategies on adjacent arterials. Thus, a good design of the metering algorithms and its proper assessment must take the entire freeway corridor into account.

On the other hand, a major capacity expansion of a given freeway, such as the addition of a lane or the extension of the HOV facility, will significantly alter trip patterns. That is, the capacity expansion will have network-wide impact, which cannot be reliably assessed by studying the freeway alone.

Thus, for the Action Plans of traffic control, incident management, traveler information systems and demand management that TOPl seeks to assess, a corridor is the appropriate unit of analysis. It may be useful to view TOPl as tools for planning *corridor management*.

Figure 1.2 summarizes the expected result of the TMS Master Plan implementation with respect to congestion relief and indicates the niche for TOPl.

The rest of the dissertation is organized as follows.

Chapter 2.

We start by explaining the difference between micro and macro approaches to traffic modeling and giving basic definitions relevant for macro modeling. In the literature review we go over continuous traffic flow models, their discretizations (we show how the Godunov scheme is applied to the LWR model) and discuss the related control problem. Then we describe

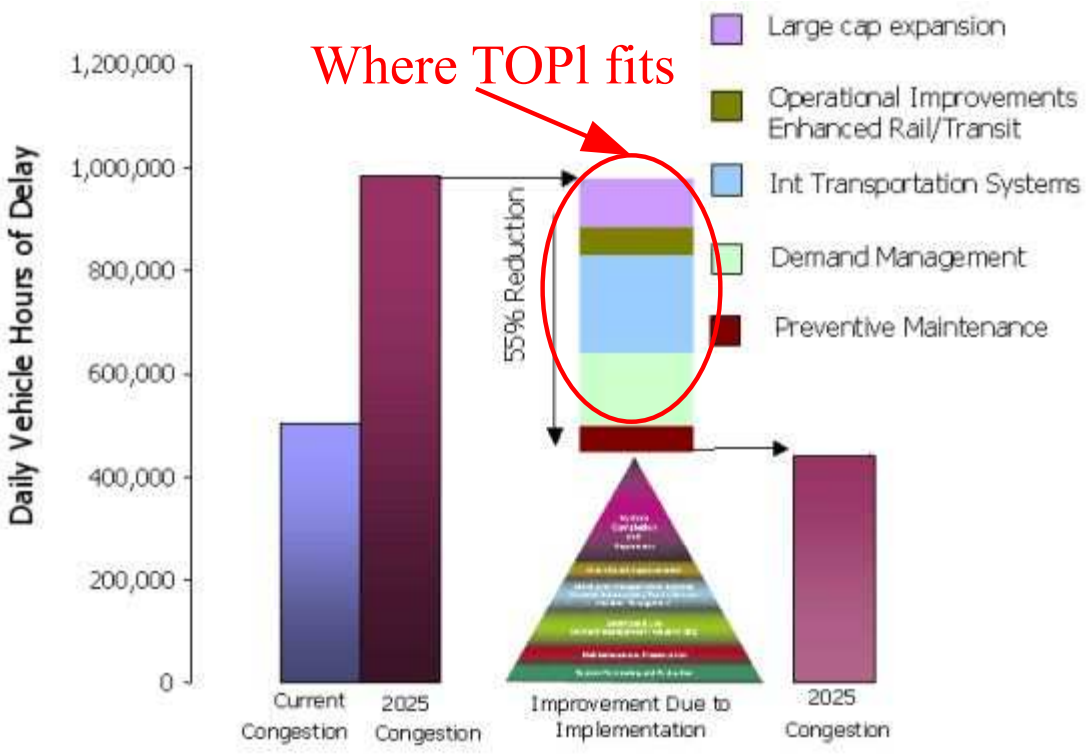


Figure 1.2: Impact of the TMS Master Plan on congestion relief.

the work on macro modeling of arterials. Finally, we discuss available software packages that could be used for the purposes of the TOPl project.

Chapter 3.

In TOPl, our traffic flow model of choice is the cell transmission model (CTM). While simple, it adequatly describes the traffic flow on freeways, as the simulation results match well the measurement data provided by PeMS.

In this chapter we first describe the CTM model. Then, studying it as a nonlinear dynamical system, we characterize its equilibrium points. It turns out that the structure of the CTM equilibria can be formalized. Moreover, if the demand is strictly below capacity, there is a

unique, globally asymptotically stable equilibrium point. Once the equilibrium structure is established, we study the dynamics of the CTM system showing that this system is monotone and that all trajectories converge. Finally we discuss the implications of these findings for ramp metering.

Chapter 4.

Having found no appropriate software implementation of a first order traffic flow model whose input could be automatically generated and whose output could be seamlessly compared with PeMS data, the TOPl group developed CTMSIM, its own interactive freeway traffic simulator for MATLAB. CTMSIM employs the CTM model, allows plugging in user-defined ramp flow and ramp queue controllers. Its output results can be directly compared with PeMS data, and it can operate in both graphical (interactive) mode and command line (batch) mode.

In this chapter, we provide a thorough description of CTMSIM. We explain its computational model, present its user interface, show how to plug in a user-defined ramp controller, how to build a freeway configuration and how to display the simulation results. Here we also present a case study of the I210-West freeway. This chapter can be viewed as a manual for CTMSIM.

Chapter 5.

The ultimate goal of TOPl is to create an instrument for corridor management. Although CTMSIM is a simple and convenient tool for freeway modeling, it lacks capabilities for corridor traffic simulation and analysis. Hence, the TOPl group started developing Aurora, a framework for simulation and analysis of infrastructure networks, whose implementation

is done in Java.

In this chapter, we provide a description of Aurora. We list the design goals, introduce basic objects used to construct road networks and event objects needed to create and run scenarios, explain the computation algorithm, and describe the organization of configuration files. Then we present the current version of the user interface. Finally, we discuss the goals of Aurora development and their priorities. This chapter can be viewed as an Aurora technical report.

Chapter 6.

In this chapter we state the problems to be solved and tasks to be accomplished within TOPl project or as a side product of TOPl activities.

Chapter 2

Review of Previous Work

2.1 Preliminaries

There exist two fundamentally different approaches to traffic modeling. The *microscopic* approach seeks to reproduce the behavior of an individual vehicle, as its driver responds to its environment by adjusting its speed and lane. Microscopic models typically involve variables such as vehicle position, speed and headway. The *macroscopic* approach ignores the dynamics of the individual vehicle and instead attempts to replicate the aggregate response of a large number of cars. These models represent traffic as a compressible fluid in terms of flow, density and speed. Traffic engineering has benefitted immensely from macroscopic models. They are widely used in the design of freeway facilities and they are present in nearly all model-based ramp metering designs. Because of its emphasis on quick and quantative assessment, TOPl's tools and procedures are based on macroscopic models that are easier

to assemble, calibrate, and automate, as compared with their microscopic counterparts[1].

The Highway Capacity Manual 2000 [101] provides the following definitions of the basic quantities. Symbols x and t represent position (measured in the direction of traffic flow) and time.

Speed $v(x,t)$ is a rate of motion expressed as distance per unit of time. Depending on how it is measured, it is referred to as either *space mean speed* or *time mean speed*. Space mean speed is computed by dividing the length of a road by the average time it takes for vehicles to traverse it. Time mean speed is the average speed of vehicles observed passing a given point. The latter is easier to measure in the field, as it can be obtained directly from conventional sensing devices.

Free flow speed is the average speed of traffic measured under conditions of low volume, when vehicles can move freely at their desired speed.

Flow $f(x,t)$ is the total number of vehicles that pass by the point x during a given time interval containing t, divided by the length of the time interval. It is usually expressed as an hourly rate, and is easily measured with road sensors.

Density $\rho(x,t)$ is the number of vehicles occupying a length of road about point x at time instant t. Its measurement is difficult because it requires the observation of a stretch of road. Instead, it is often approximated from measurements of flow and speed as

$$\rho(x,t) = \frac{f(x,t)}{v(x,t)}. \tag{2.1}$$

[1]Carrying out micro simulations for all plausible Action Plans is not practical. For example, a study uncovered more than 500 bottlenecks [71], the congestion caused by which could be mitigated by ramp metering. It is not possible to study all these opportunities by micro simulations.

Demand is the number of vehicles that desire to use a given facility during a specified time period.

Capacity is the maximal hourly rate at which vehicles reasonably can be expected to traverse a point or a uniform section of a lane or roadway during given time period under prevailing roadway, traffic and control conditions.

Bottleneck is defined as any road element where demand exceeds capacity. Freeway bottlenecks sometimes appear near heavy on-ramps, where a localized increase in demand is combined with a decrease in capacity due to lane changing.

One of the early attempts to correlate freeway speed, density and flow was by Greenshields in 1934 [53]. He used photographic images to estimate aggregate vehicular speeds and densities on a straight two-lane roadway, and found that they could be reasonably well approximated by a straight line. Using (2.1) he derived parabolic relationship between flow and density as shown in Figure 2.1. Function $f = \Phi(\rho)$ is known as the *fundamental diagram*. Later

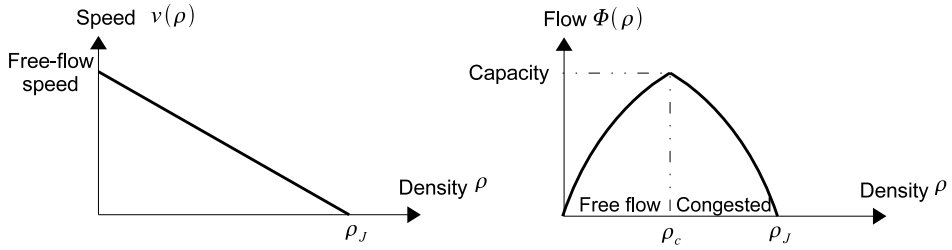


Figure 2.1: Greenshields' speed and flow relations.

researchers have suggested alternative shapes that provide a better fit to the measured data (see Figure 2.2). All of them share the following characteristics:

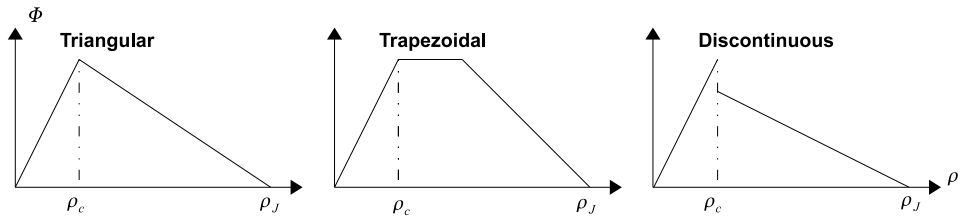


Figure 2.2: Alternative shapes for the $\Phi(\rho)$ function.

1. $\Phi(0) = \Phi(\rho_J) = 0$, where ρ_J is called *jam density*.

2. Continuous portions of $\Phi(\rho)$ are concave.

3. *Critical density* ρ_c can be defined where the maximum flow is attained. Then, $\Phi'(\rho) \geq 0$ for $\rho < \rho_c$ and $\Phi'(\rho) \leq 0$ for $\rho > \rho_c$

Critical density ρ_c splits the fundamental diagram into two regimes: *free flow* ($\rho \leq \rho_c$) and *congestion* ($\rho > \rho_c$) (see Figure 2.1). Measurements on the free flow side are usually well represented by a straight line, whereas measurements in congestion tend to be more scattered.

2.2 Theoretical Background

2.2.1 Continuous Time Models

First order LWR model.

The simplest continuous macroscopic model is the scalar one proposed by Lighthill and Whitham [77], and by Richards [93]. Hence, this model is called LWR. Lighthill and

Whitham in 1955 were the first to pose a macroscopic dynamic model of traffic using Greenshields' hypothesis of a static flow/density relationship. LWR is based on conservation of cars and is described by a single nonlinear hyperbolic equation, also known as *conservation law*:

$$\rho_t + (\Phi(\rho))_x = 0, \tag{2.2}$$

where function Φ is the flow. In this model, the average speed v is a function that depends only on density. The relation $\Phi(\rho) = \rho v(\rho)$ is a fundamental diagram and is classically assumed to be concave (does not need to be parabola, see Figure 2.3). It is defined for $\rho \in [0, \rho_J]$, where ρ_J is the jam density and corresponds to the density at which traffic stops. The density ρ_c for which the flow reaches maximum (the road operates at capacity), is the critical density. Traffic speed $v \geq \Phi(\rho_c)/\rho_c$ is called *free flow speed*. When the density exceeds critical, the road becomes congested: the traffic speed falls below free flow, $v < \Phi(\rho_c)/\rho_c$.

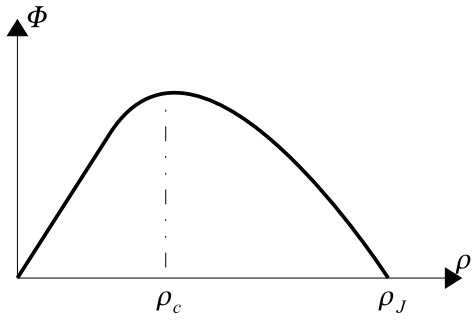


Figure 2.3: Fundamental diagram.

To include on- and off-ramps into the LWR model (Figure 2.4), we rewrite (2.2) in integral

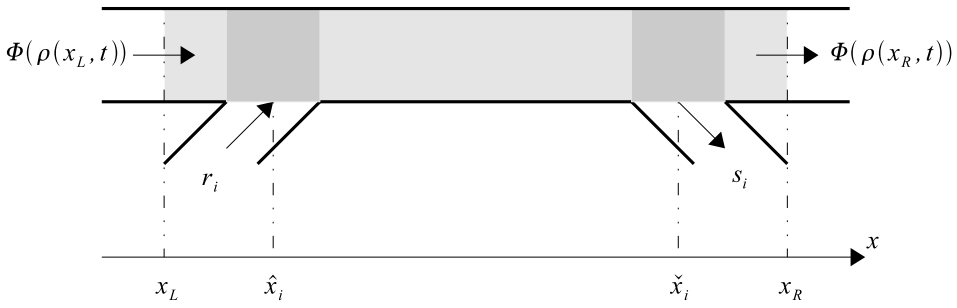


Figure 2.4: Freeway section with on- and off-ramp.

form and account for the on-ramp flow r_i and off-ramp flow s_i:

$$\frac{d}{dt}\int_{x_L}^{x_R}\rho(x,t)dx = \Phi(\rho(x_L,t)) - \Phi(\rho(x_R,t)) + r_i - s_i,$$

which can be once again rewritten as

$$\int_{x_L}^{x_R}\rho_t(x,t)dx = \int_{x_L}^{x_R}\left((\Phi(\rho(x,t)))_x + \delta(x-\hat{x}_i)r_i(t) - \delta(x-\check{x}_i)s_i(t)\right)dx, \tag{2.3}$$

where $\delta(x)$ is a Dirac delta function.

For multiple on-ramps ($N_{on} \geq 1$) and off-ramps ($N_{off} \geq 1$) equation (2.3) generalizes to

$$\rho_t(x,t) + (\Phi(\rho(x,t)))_x = \sum_{i=1}^{N_{on}}\delta(x-\hat{x}_i)r_i(t) - \sum_{i=1}^{N_{off}}\delta(x-\check{x}_i)s_i(t). \tag{2.4}$$

Clearly, in the absence of ramps, equation (2.4) becomes (2.2).

Second order Payne-Whitham model.

The assumption of the LWR model about the average speed v depending only on density is not valid for certain traffic situations, such as capacity drop, an empirical feature of freeway traffic, which is the difference of the maximum observed flow and the flow exiting the bottleneck when there is a congestion upstream of the bottleneck. Hence, Payne [91]

introduced an additional equation for v, producing the following system of equations:

$$\rho_t + (\rho v)_x = 0, \tag{2.5}$$

$$v_t + vv_x + \frac{1}{\rho}(A_e(\rho))_x = \frac{1}{\tau}(v_e(\rho) - v), \tag{2.6}$$

where $v_e(\rho)$ is the equilibrium value for the speed, $\frac{1}{\rho}(A_e(\rho))_x$ is called anticipation term, and $\frac{1}{\tau}(v_e(\rho)-v)$ is called relaxation term of v within a certain time $\tau > 0$ towards its equilibrium value $v_e(\rho)$. Equation (2.5) is the *continuum* equation and (2.6) is the *acceleration* equation.

Whitham himself proposed a generalization of the LWR model [108] adding the following acceleration equation:

$$v_t + vv_x + \frac{D}{\rho}\rho_x = \frac{1}{\tau}(v_e(\rho) - v)$$

with some constant $D > 0$. Clearly, this last equation is a special case of (2.6) with $A_e(\rho) = D\rho$. Therefore, a model with system (2.5)-(2.6) is called Payne-Whitham model.

Payne-Whitham model develops analogy with gas dynamics. As in the gas dynamic case, the term vv_x is called the *convection* term and describes a motion of the speed profile. The anticipation term reflects the reaction of identical drivers to the surrounding traffic situation. The relaxation term describes the adaptation of the average speed v to the equilibrium speed $v_e(\rho)$.

Severe drawbacks of the Payne-Whitham model were pointed out by Daganzo in [41]. In particular, he showed that unlike in fluids where a particle's behavior depends on the particles in front as well as on the particles behind, a car particle is only affected by the particles in front; and that in Payne-Whitham model cars are allowed to travel with negative speeds, which is unrealistic.

Second order ARZ model.

In an effort to rehabilitate the second order models and address the issues pointed out by Daganzo, Aw and Rascle in [18] and independently Zhang in [110] came up with the following second order, now known as Aw-Rascle-Zhang (ARZ), model:

$$\rho_t + (\rho v)_x = 0, \tag{2.7}$$

$$(v + p(\rho))_t + v(v + p(\rho))_x = 0, \tag{2.8}$$

where $p = p(\rho)$ is a "pressure" term, an increasing function of density. Aw and Rascle considered

$$p(\rho) = \rho^\gamma, \;\; \gamma > 0,$$

and Zhang used

$$p(\rho) = -v_e(\rho) = -\frac{\Phi(\rho)}{\rho}.$$

Third order model.

The first *third order* model was proposed by Helbing in [54]. He considers not only equations for density ρ and velocity v, but also for the velocity variance θ. The exact model proposed by Helbing is

$$\rho_t + (\rho v)_x = 0, \tag{2.9}$$

$$v_t + v v_x + \frac{1}{\rho}(\rho\theta)_x = \frac{1}{\tau}(v_e(\rho) - v) + \frac{\mu}{\rho} v_{xx}, \tag{2.10}$$

$$\theta_t + v\theta_x + 2\theta v_x = 2\frac{\mu}{\rho}(v_x)^2 + \frac{k}{\rho}\theta_{xx} + \frac{2}{\tau}(\theta_e(\rho) - \theta), \tag{2.11}$$

where θ_e and v_e are given smooth functions of density ρ, and μ, k, τ are nonnegative constants. The term $-\frac{k}{\rho}\theta_{xx}$ results from the finite reaction and braking time that causes a

delayed adaption of velocity to the respective traffic situation. The term $\frac{2}{\tau}(\theta_e - \theta)$ results on one hand from the drivers' attempt to drive with desired speed v_e, and on the other hand - from the drivers' interactions, i.e. from deceleration maneuvers in cases when a fast vehicle cannot overtake a slow one. The desired velocity v_e varies from one driver to another. Therefore, even for small values of density ρ a finite velocity variance $\theta_e(\rho)$ of the vehicles is expected. By analogy with gas dynamics, it is said that model (2.9)-(2.11) is of Euler type, if coefficients μ and k both equal 0. Otherwise, this model is said to be of Navier-Stokes type. Stability analysis and numerical simulations for the model (2.9)-(2.11) can be found in [55]. As shown in [54], this model is adequate to describe the stop-and-go traffic and the velocity variance θ can be used to predict traffic jams as its value grows immediately before the congestion starts developing.

Hyperbolic phase transition model.

A *hyperbolic phase transition* model for traffic was introduced by Colombo in [35]. He considers two phases corresponding to free and congested flows. In free flow phase the LWR equation (2.2) holds. However, when the density goes beyond critical, the assumption that the speed v is a function only of ρ is no longer valid, and the density-flow points are scattered in a two-dimensional region (see Figure 2.5). Thus, the model is described by its free flow phase

$$(\rho, q) \in \Omega_f, \tag{2.12}$$

$$\rho_t + (\rho v)_x = 0, \tag{2.13}$$

$$v = (1 - \frac{\rho}{\rho_J})v_{\max} \tag{2.14}$$

and by its congested phase

$$(\rho, q) \in \Omega_c, \tag{2.15}$$

$$\rho_t + (\rho v)_x = 0, \tag{2.16}$$

$$q_t + ((q - Q)v)_x = 0, \tag{2.17}$$

$$v = (1 - \frac{\rho}{\rho_J})\frac{q}{\rho} \tag{2.18}$$

Here Ω_f and Ω_c denote free flow and congested domains, the weighted flow q is a variable

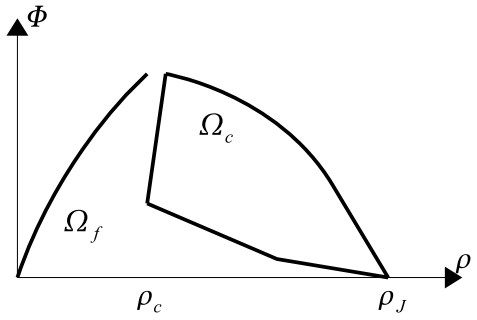


Figure 2.5: Fundamental diagram for the hyperbolic phase transition model.

originally motivated by the linear momentum in gas dynamics, the treshold parameter Q distinguishes between possible behaviors of the flow (see [36]), and $v_{\max}$ is the maximum speed.

2.2.2 Discretization

As conservation laws can have discontinuous solutions, they cannot be integreated numerically by standard methods such as finite differences or finite elements that create instabilities and wrong shock speeds [75]. Among the numerical schemes for scalar and systems of conservation laws [75, 49] the Godunov scheme [50] is widely used. It is first order, correctly

predicts shock propagations, is free of oscillating behavior and has physical interpretation. In this approach the time is discretized into intervals $[k\Delta t, (k+1)\Delta t]$. The computational domain is divided into cells[2], and at time $k\Delta t$, the solution ρ of (2.2) is approximated by a piecewise constant function $\tilde{\rho}$ (see Figure 2.6) defined as

$$\tilde{\rho}(x, k\Delta t) = \rho_i^k, \quad \forall i, \ \forall x \in [x_{i-1}, x_i].$$

The computation of the approximation $\tilde{\rho}(\cdot, (k+1)\Delta t)$ using the approximation $\tilde{\rho}(\cdot, k\Delta t)$

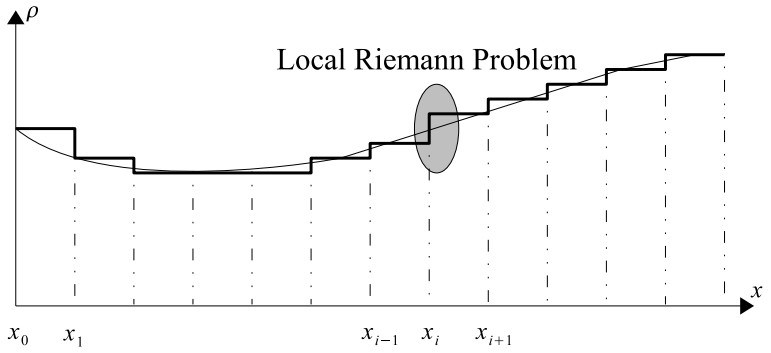


Figure 2.6: Piecewise constant approximation of the state.

requires two steps.

1. Compute exact solution of (2.2) given the initial condition

$$\rho(x, k\Delta t) = \tilde{\rho}(x, k\Delta t) = \rho_i^k, \quad \forall i, \ \forall x \in [x_{i-1}, x_i]. \tag{2.19}$$

2. Take the average of $\rho(\cdot, (k+1)\Delta t)$ over every cell $[x_{i-1}, x_i]$:

$$\rho_i^{k+1} = \frac{1}{\Delta x_i} \int_{x_{i-1}}^{x_i} \rho(y, (k+1)\Delta t) dy.$$

[2]Here and throughout the dissertation cell numbers increase in the direction of traffic flow: cell i is upstream of cell $i+1$.

These two steps can be simplified as follows:

$$\rho_i^{k+1} = \rho_i^k + \frac{\Delta t}{\Delta x_i}(f_{i-1}^k - f_i^k) \tag{2.20}$$

with

$$f_i^k = \frac{1}{\Delta t}\int_{k\Delta t}^{(k+1)\Delta t} \Phi(\rho(x_i, s))ds \tag{2.21}$$

being the average flow crossing x_i from cell i to cell $i+1$ during the time interval $[k\Delta t, (k+1)\Delta t]$. Finally, since function Φ is concave, expression (2.21) can be replaced by

$$f_i^k = \begin{cases} \min_{\rho_i^k \le \rho \le \rho_{i+1}^k} \Phi(\rho), & \text{if } \rho_i^k \le \rho_{i+1}^k, \\ \max_{\rho_{i+1}^k \le \rho \le \rho_i^k} \Phi(\rho), & \text{if } \rho_i^k \ge \rho_{i+1}^k. \end{cases} \tag{2.22}$$

In summary, the Godunov scheme leads to a piecewise approximation of the state (density) ρ at each time step, whose evolution can be computed for small time intervals if we know the solutions of initial value problems with Heaviside initial conditions

$$\rho(x) = \begin{cases} \rho^-, & x < 0 \\ \\ \rho^+, & x > 0 \end{cases}. \tag{2.23}$$

Such initial value problem is an abstraction of the problem (2.2), (2.19), and is called a *Riemann problem*. It can be solved analytically for scalar conservation laws [75], and in the system case, when there is no closed form solution, an approximate solver such as the Roe average method [75, 49] can be used. The Godunov scheme, consisting in solving a succession of local Riemann problems, is an effective method for simulating macroscopic traffic models.

An example of Godunov scheme in action is the STRADA model [27], which is based on two extensions of the basic LWR: a simple first in, first out model for the density dynamics in

cells and a macroscopic model for the dynamics of traffic through intersections, consistent with the boundary conditions resulting from the LWR.

Details of the LWR discretization via Godunov scheme can be found in [74].

The *cell transmission model* (CTM) proposed in [39] is another special case of the Godunov difference scheme where the fundamental diagram has triangular form with maximal flow F, slope $v > 0$ for the free flow speed and slope $-w < 0$ for the congestion wave speed (see figure 2.7). In this framework, the Godunov scheme becomes

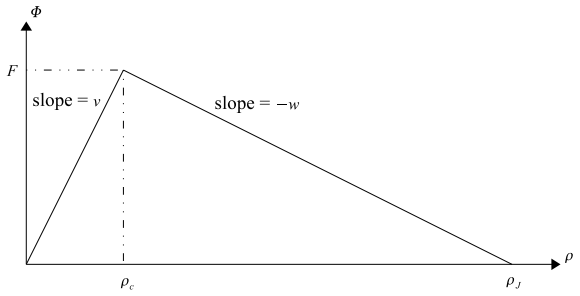


Figure 2.7: The fundamental diagram for CTM is characterized by the maximum flow F and speeds v, w.

$$\rho_i(t+1) = \rho_i(t) + \frac{\Delta t}{\Delta x_i}(f_{i-1}(t) - f_i(t)),$$

where Δt is the sampling period, Δx_i is the length of the ith cell, and f_i, the flow from cell i to cell $i+1$, is given by

$$f_i(t) = \min\{v\rho_i(t), w(\rho_J - \rho_i(t)), F\}.$$

Consequently, cell i can operate in one of two modes: *free flow mode* if $f_{i-1} = \min\{v\rho_{i-1}, F\}$, or *congested mode* if $f_{i-1} = w(\rho_J - \rho_i)$.

Variations on the CTM theme can be found in literature. In [51] the *asymmetric cell transmission model* (ACTM) is presented. This work also describes a convex optimization problem whose solution is an optimal ramp metering strategy.

The linear hybrid system approach called *switching mode model* (SMM) based on CTM is introduced in [86].

The Godunov method can be effectively applied to the second order ARZ models. The details of discretization for various types of fundamental diagrams are described in [59].

By spatial discretization of the Payne-Whitham model (2.5)-(2.6), Smulders [97] defined a macroscopic flow model that is continuous in time and discrete in space. Furthermore, using the theory of martingales, he introduced a stochastic component based on a counting process. The flow f is determined by a convex sum approximation of the average density ρ and the speed v of the adjacent cells i and $i+1$: The density dynamics is given by

$$d\rho_i = (f_{i-1} - f_i)\frac{dt}{\Delta x} + \frac{d\xi_{i-1} - d\xi_i}{\Delta x}, \tag{2.24}$$

where ξ_i reflects the stochastic departure process of vehicles from segment i, and the outflow of cell i is determined by a convex approximation of the average density ρ and the speed v of the adjacent cells i and $i+1$:

$$f_i = (\alpha\rho_i + (1-\alpha)\rho_{i+1})(\alpha v_i + (1-\alpha)v_{i+1}), \tag{2.25}$$

with $0 \leq \alpha \leq 1$.

The speed dynamics is given by

$$dv_i = \frac{1}{\Delta x}v_{i-1}(v_{i-1} - v_i)dt + \frac{1}{\tau}(v_e(\rho_i) - v_i)dt - \frac{D}{\Delta x}\frac{\rho_{i+1} - \rho_i}{\rho_i + c}, \tag{2.26}$$

with $c > 0$ being a constant.

With respect to the conservation equation (2.24) this approximation yields the conservation of vehicles. However, unless $\alpha = 1$, vehicles may flow out of an empty downstream cell possibly generating negative density values. Moreover, if the downstream cell is congested ($\rho_{i+1} = \rho_J$ and $v_{i+1} = 0$), vehicles will still flow out of the upstream cell into the saturated downstream cell.

Examples of discrete Payne-Whitham type models can be found in [92, 102, 89, 66, 84, 80].

2.2.3 Control of Conservation Laws

In the context of traffic flow applications, the goal of control is to improve the system efficiency by regulating the number of vehicles allowed to enter the freeway. Two fundamental performance measures are used to assess the system efficiency: the *total travel distance* (TTD) and the *total travel time* (TTT). TTD is defined as the sum of distances traveled by all vehicles of the system over a given time period. Equivalently, it is a product of the average trip length and the total number of vehicles, which can be computed as the integral of flow over time and space:

$$\text{TTD} = \int_X \int_T f(x,t)dtdx. \tag{2.27}$$

TTT is the sum of all trip times incurred by vehicles during a given time period, or the number of vehicles multiplied by the average trip time, which is computed as the integral of density over time and space:

$$\text{TTT} = \int_X \int_T \rho(x,t)dtdx. \tag{2.28}$$

The goal of control is regulating the number of vehicles entering the freeway either to maximize TTD or minimize TTT.

The optimal control theory of partial differential equations was initiated in the early 70's by Lions [79]. The proposed approach consists in computing the necessary conditions of optimality in the form of the system equation, an adjoint equation of the same kind and a vanishing first variation condition. This analytic approach that was successfully applied to linear elliptic, parabolic and second order hyperbolic equations can be extended to nonlinear systems using gradient-based recursive algorithms. This method is widely used: in airfoil design [60, 62, 61]; fluid steering [23, 56, 34, 47]; gas steering [48]; control of water wave [94, 32]; air traffic control [20].

Very few attempts have been made to stabilize conservation laws using feedback control. In [43, 37], the authors propose a feedback controller for open channels but consider only smooth solutions and no shock waves. Krstic [67] proposed a feedback design for the Burgers equation with small viscosity parameter. Unfortunately, as the control law is inversely proportional to this parameter, the controller blows up in the nonviscous case. Successful control design using a finite dimensional discretization has been reported in [19] for parabolic partial differential equations. The main difficulty in applying this method to hyperbolic conservation laws is that the classical finite difference scheme cannot be used for this class of equations due to possible presence of shock waves.

The problem of control of a system of conservation laws although addressed in the literature, remains difficult. A way around the problem, is to discretize the system first, then solve the control problem for the resulting dynamical system. An example of such approach is

the multirate linear quadratic control with integral action (LQI) [98]. Jacquet [59] shows how Godunov discretization can be put in the form of a piecewise affine system if the fundamental diagram is approximated by a piecewise affine function, and suggests that constructive controller design methods proposed in [64, 38, 24] can be used to compute a set of static feedback gains for a switched controller.

2.2.4 Arterial Models

Ziliaskopoulos and Lee adapt CTM [40] for arterial modeling [113]. The cell length is generally much shorter for arterials than for freeways, hence the sampling period Δt must be small enough to ensure

$$v\Delta t < l,$$

where v is the free flow speed (Figure 2.7) and l is the cell length.

Signalized intersections are modeled using diverging and merging cells[3], and the signal phasing (red and green). The flow of the diverging cells is computed according to the CTM during the green phase and is set to zero during the red phase. In [112] CTM is used to formulate the system optimum dynamic traffic assignment problem[4] as a linear programming (LP) problem.

In [81], Lo transforms CTM into a set of mixed-integer constraints and casts the dynamic signal-control problem[5] to a mixed-integer linear program. As a dynamic platform, this

[3]Diverging are the cells with one predecessor and two or more successors. Merging are the cells with two or more predecessors and one successor.

[4]See Section 5.4 for definition of the system optimum dynamic traffic assignment problem.

[5]The problem of red/green signal phase assignment so as to minimize total travel time, number of slowed down vehicles (vehicles with speed below free flow speed), or maximize total out-flow of the system.

formulation is flexible in dealing with dynamic timing plans and traffic patterns. It derives dynamic as well as fixed timing plans and addresses preexisting traffic conditions and time dependent demand patterns. Dynamic intersection signal control optimization (DISCO) that works with time-variant traffic patterns and derives signal timing plans is introduced in [82]. The authors compare DISCO with the platoon based TRANSYT model [2] and conclude that timing plans generated by DISCO outperform those generated by TRANSYT by as much as 33% in delay reduction under a variety of demand patterns.

Feldman and Maher [45] investigate CTM applicability to the network of signalized arterials and compare it with the TRANSYT model [2]. Modeling the arterial with a pair of traffic signals with both CTM and TRANSYT, the authors conclude that CTM yields similar or better results than TRANSYT does.

Amasri and Friedrich [14] also apply CTM to urban arterials and compare it with queueing models. They use genetic algorithm (GA) to find optimal signal timing plan having CTM as an underlying traffic flow model.

Alecsandru in [13] suggests modifications to CTM that include some microscopic features such as disaggregating the traffic flow by lanes and explicitely modeling the effects of individual lane-changing maneuvers; replacing some of the original parameters in the analyzed network with stochastic variables to capture the effect of the random driving behavior; and changes to the model equations that allow to keep track of different vehicle types. He also compares this modified CTM with CORSIM [58] microsimulation, and shows that the simulation outputs (traffic density and total network travel time) of these two models match well.

Nie [88] presents a polymorphic dynamic network loading (PDNL) framework for modeling road networks and solving the dynamic network loading problems[6]. PDNL employs notions of links and nodes allowing different macroscopic traffic flow models to run on links while treating nodes as points of merge, diverge or general intersections, signalized or not.

Skabardonis and Geroliminis [95] propose an analytical model for travel time estimation on arterials. Their model is based on CTM, describes the spatial and temporal queuing at traffic signals and explicitely considers the signal coordination in estimating traffic arrivals at intersections. It estimates the travel time over an arterial link as the sum of free flow time and the delay at traffic signal.

In these works authors do not discuss computational complexity of the proposed models. The question how the size of a road network affects the efficiency of the proposed algorithms remains open.

2.3 Software Tools

FREQ.

We start the description of macrosimulation software with *FREQ* [3]. It was developed in the University of California, Berkeley, since 1968. FREQ employs a first order, LWR, model and implements the recommendations of the HCM [101]. Current version of the software is 12. It includes interactive graphical user interface with input checking and pre-selected

[6]Problems that aim at obtaining the link cumulative arrival/departure curves (hence time-dependent link/path travel times) corresponding to a given set of temporal path flow rates on a congested network and over a fixed time period.

default values, graphical representation of simulation results, user-defined output options including traffic performance contour maps. The two models contained within FREQ12 are: FREQ12PE, an entry control macroscopic model for analyzing ramp metering; and FREQ12PL, a freeway macroscopic model for analyzing HOV facilities[7].

The FREQ system of models enables the user to analyze geometric design improvements, an on-freeway HOV facility, normal and priority entry control, or time-varying capacity reduction situations such as reconstruction activities or freeway incidents. The analysis emphasizes traffic simulation, traveler responses, and measures of effectiveness.

All this said about FREQ12, we must acknowledge that this tool is not user-friendly and can hardly be used for practical purposes. The main drawback is the lack of documentation. Only feature changes from version to version are documented, leaving the core of the simulator to be treated as black box. Despite some context help provided by the application, only a person thoroughly familiar with HCM [101] may feel comfortable adjusting parameters and fully using features of FREQ. The second major hindrance to using FREQ is the fact that configuration files for the simulator must be created from within the application itself. The format of the binary configuration file is proprietary, and thus, the configuration file cannot be automatically generated making it necessary for the user to manually fill in the parameter values for every cell and ramp. The same is true for demand and off-ramp flow profiles—they have to be manually typed in from within the FREQ application. Another serious limitation of FREQ is its inability to model ramp control. No built-in ramp controllers are provided and it is impossible to plug in user-defined controllers. This makes FREQ unsuitable for

[7]We could not find any documentation describing these models.

modeling traffic responsive ramp metering strategies. Finally, the graphical output can be printed but not saved making report generation using FREQ simulation results inconvenient.

The popularity of FREQ can be explained by historical reason. It was used by transportation engineers for many years for lack of any better macroscopic simulation tool. The developers of FREQ, however, never addressed issues of proper software design adding features incrementally from version to version and not looking at the application as a whole. The result is poor user interface and lack of documentation.

NETCELL.

NETCELL [30] is another macrosimulator developed at UC Berkeley. It implements the CTM for networks [40]. The package consists of two applications: the simulator and the plotter. The simulator produces text output, which can then be parsed by the plotter which displays the simulation results along with the network configuration. NETCELL is free and can be downloaded from [4]. As far as we know, NETCELL is currently not supported and is hardly used by anyone because of the lack of proper user interface.

METACOR.

METACOR [44] is the first macroscopic simulation tool for corridor traffic, i.e. when freeways and arterials are modeled together in one network. It emerged as a fusion of *METANET* [85] (for freeways) and *SSMT* [73], a macroscopic model for urban networks. METACOR uses a discrete version of the Payne-Whitham second order model. It also includes control and dynamic traffic assignment modules to simulate ramp metering strategies and route information/guidance via changeable message signs. Currently, METACOR is being developed independently by the Technical University of Crete (TUC) and the Institut National

de Recherche sur les Transports et leur Sécurité (INRETS). Individual copies of the former version can be obtained from M. Papageorgiou. He warns, though, that the user interface is rather primitive and there is no documentation in English. The latter version, known as *PX-Metacor* [5] is supposed to be commercially available. The advertisement says that additionally to Payne-Whitham, it supports LWR and ARZ models as well as micro- and hybrid (macro + micro) simulation, and has quality graphical user interface. We were unable to get hold of PX-Metacor. It is still under development and has not been released at the time of this writing.

SATURN.

SATURN [103, 104] was developed at the University of Leeds to evaluate traffic management schemes on arterial networks. It contains an equilibrium traffic assignment algorithm based on macroscopic flow relationships, where the travel time is an increasing function of flow. It has been widely used to evaluate changes in circulation (one-way streets, pedestrianization schemes) and other traffic management schemes. It has also been used to evaluate the effectiveness of route guidance systems and road pricing studies. The major limitation of SATURN is that it was developed specifically for arterials and is unsuitable for freeway modeling. SATURN has developed into a commercial application [6].

CORFLO.

CORFLO [76] contains the *FREFLO* [92] model for freeways and the *NETFLO1*, *NETFLO2* models for arterials. The interface of adjoining subnetworks is accomplished by defining interface nodes representing points at which vehicles leave one subnetwork and enter another. Associated with each subnetwork is a vehicle holding area where exiting vehicles are held

until the next subnetwork can receive them. Traffic may be assigned to the different subnetworks using the *TRAFFIC* assignment model (a static equilibrium assignment algorithm). FREFLO is based on the Payne-Whitham model. Unfortunately, it had serious errors in implementation, which caused unrealistic simulation results and were never fixed. It can handle different vehicle classes (busses, carpools), HOV facilities, and incidents on the freeway, but it cannot model ramp operations. NETFLO1 is a microscopic event scanning simulator. NETFLO2 models traffic using flow profiles similar to the TRANSYT model [2]. Unlike TRANSYT, however, it can simulate signals with different cycle lengths and queue spillbacks. There is little information published on the development and application of the NETFLO models. CORFLO has been designed to evaluate freeway and arterial design and control modifications, impacts of incidents and diversion policies. As far as we know, CORFLO is currently obsolete.

DYNASMART.

DYNASMART [63] was developed at the University of Texas at Austin as both simulation and assignment tool. Traffic flow is simulated using the Payne-Whitham model. It can simulate traffic signals, ramp meters and incidents. It calculates optimal travel paths based on the simulated travel times, and simulates the movements and routing decisions by individual drivers equipped with in-vehicle information systems (update of information and desire to switch based on thresholds). DYNASMART is available as a research tool and is now supported by the University of Maryland [7]. Two versions of DYNASMART are available: DYNASMART-X for real-time analysis, and DYNASMART-P for planning.

Chapter 3

CTM: Qualitative Theory

This chapter largely reproduces [52]. It studies the *cell transmission model* (CTM) [39, 40] as a dynamical system. The freeway is divided into N cells, indexed $1, \cdots, N$. Cell i is characterized by a single state variable, its density n_i, so the state of the freeway is the N-dimensional vector $n = (n_1, \cdots, n_N)$. Vehicle movement in a cell is governed by the familiar triangular fundamental diagram. If the density is below critical, vehicles move at free flow speed; if it is above critical, the cell is congested, speed is lower, and flow from the immediately upstream cell is constrained. Thus the state of a freeway obeys a N-dimensional nonlinear difference equation. When the exogenous demand pattern of on-ramp and off-ramp flows is constant, the difference equation is time-invariant.

CTM is popular for its flexible use in macroscopic simulation. Compared with microscopic simulators, it requires negligible computational effort. It can be extended to road networks ([27]) and urban roads with signalized intersections ([81, 14]). It appears well-suited for

calibration with point measurements of aggregate traffic variables that are routinely available ([78, 87]). It is straightforward to formulate questions of ramp metering ([42, 111, 51]) and dynamic traffic assignment ([28, 112]) by posing optimization problems within CTM.

Proofs for lemmas can be found in Section 3.5.

3.1 The Model

Following [51], figure 3.1 shows the freeway divided into N cells, each with one on- and one off-ramp. Vehicles move from left to right. Cell i is upstream of cell $i+1$. There are two boundary conditions. Free flow prevails downstream of cell N; upstream of the freeway is an "on-ramp" with an inflow of r_0. The flow accepted by cells 1 is $f_0(k)$ vehicles per period or time step k. The cumulative difference leads to a queue of size $n_0(k)$ in period k.

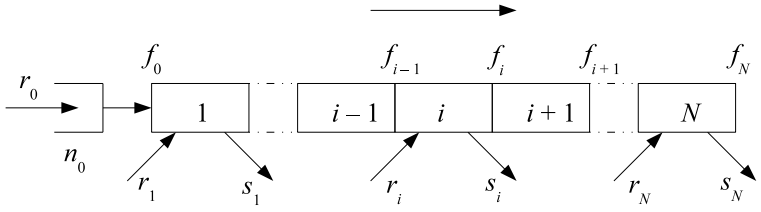


Figure 3.1: The freeway has N cells. Each cell has one on- and one off-ramp.

Table 3.1 lists the model variables and parameters with plausible values. The length of all cells is normalized to 1 by absorbing differences in length in the speeds v_i, w_i. To be physically meaningful one must have $0 < v_i, w_i < 1$. The parameter values in the table correspond to the fundamental diagram of Figure 3.2. Its triangular form incorporates the

Symbol	Name	Value	Unit
	cell length	1	miles
	period (time step)	0.5	minutes
F_i	capacity	20	veh/period
v_i	free flow speed	0.5	cell/period
w_i	congestion wave speed	0.5/3	cell/period
$\bar{n}_i$	jam density	160	veh/cell
n_i^c	critical density	40	veh/cell
β_i	split ratio	$\in [0, 1)$	dimensionless
$\bar{\beta}_i$	complementary split ratio $= 1 - \beta_i$	$\in (0, 1]$	dimensionless
γ_i	on-ramp blending factor	$\in [0, 1]$	dimensionless
k	period number	integer	dimensionless
$f_i(k)$	flow from cell i to $i+1$ in period k	variable	veh/period
$s_i(k), r_i(k)$	off-ramp, on-ramp flow in cell i in period k	variable	veh/period
$n_i(k)$	number of vehicles in cell i in period k	variable	veh/cell

Table 3.1: Model parameters and variables.

assumption that is frequently used in our analysis:

$$F_i \;=\; \bar{\beta}_i \, v_i \, n_i^c \;=\; w_i \, (\bar{n}_i - n_i^c). \tag{3.1}$$

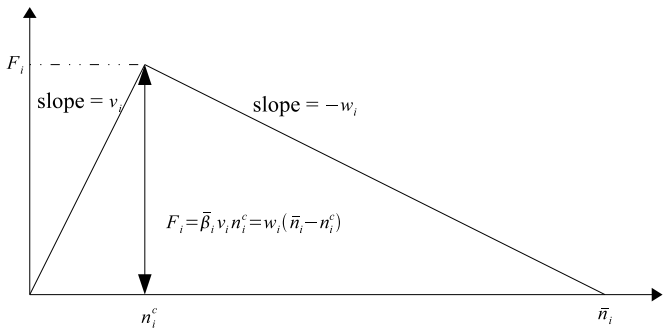


Figure 3.2: The fundamental diagram is characterized by the maximum flow F_i and speeds v_i, w_i.

Off-ramp flows are modeled as a portion $\beta_i(k)$ of the total flow leaving the cell:

$$s_i(k) = \beta_i(k)(s_i(k) + f_i(k)), \text{ or } s_i(k) = [\beta_i(k)/(1 - \beta_i(k))]f_i(k).$$

We assume constant split ratios β_i ($\beta_N = 0$). With $\bar{\beta}_i = 1 - \beta_i$, the CTM model for $k \geq 0$ is

$$n_i(k+1) = n_i(k) - f_i(k)/\bar{\beta}_i + f_{i-1}(k) + r_i(k) \quad 1 \leq i \leq N, \tag{3.2}$$

$$f_i(k) = \min\{\bar{\beta}_i\, v_i[n_i(k) + \gamma_i r_i(k)], w_{i+1}[\bar{n}_{i+1} - n_{i+1}(k) - \gamma_{i+1} r_{i+1}(k)], F_i\}, \quad 0 \leq i < N, \tag{3.3}$$

$$f_N(k) = \min\{\bar{\beta}_N\, v_N[n_N(k) + \gamma_N r_N(k)], F_N\}, \tag{3.4}$$

$$n_0(k+1) = n_0(k) - f_0(k) + r_0(k). \tag{3.5}$$

Flow conservation in cell $i > 0$ is expressed by

$$n_i(k+1) = n_i(k) - f_i(k) + f_{i-1}(k) + r_i(k) - s_i(k),$$

which is equivalent to (3.2), using $s_i(k) = \frac{\beta_i}{\bar{\beta}_i} f_i(k)$. Flow conservation at 0 is expressed by (3.5). The flow $f_i(k)$ from cell i to $i+1$ is governed by the fundamental diagram (3.3) with this interpretation: $\bar{\beta}_i\, v_i[n_i(k) + \gamma_i r_i(k)]$ is the number of vehicles that can move from cell i to $i+1$ in period k; $w_{i+1}[\bar{n}_{i+1} - n_{i+1}(k) - \gamma_{i+1} r_{i+1}(k)]$ is the number that $i+1$ can accept; and F_i is the capacity or maximum flow from cell i to $i+1$. Equation (3.4) indicates there is no congestion downstream of cell N. Lastly, it is implied that the flows $s_i(k)$ are not constrained by off-ramp capacity.

The *state* of the system is the N-dimensional vector $n(k) = (n_1(k), \cdots, n_N(k))$. The queue size $n_0(k)$ is *not* included in the state. Cell i is said to be *uncongested* in period k if $n_i(k) \leq n_i^c$, and *congested* - otherwise.

3.2 Structure of Equilibria

The parameters $\gamma_i \in [0,1]$ in (3.3), (3.4) reflect the relative position of the on-ramp in cell i ([51]). For simplicity we assume $\gamma_i = 0$, indicating that the on-ramp is at the beginning of each cell as in Figure 3.1. (However, the results below hold for a different choice of γ_i.) With $\gamma_i = 0$, equations (3.2)-(3.5) simplify:

$$n_i(k+1) = n_i(k) - f_i(k)/\bar{\beta}_i + f_{i-1}(k) + r_i(k), \quad 1 \le i \le N, \tag{3.6}$$

$$f_i(k) = f_i(n(k)) = \min\{\bar{\beta}_i\, v_i n_i(k), w_{i+1}[\bar{n}_{i+1} - n_{i+1}(k)], F_i\}, \quad 0 \le i < N, \tag{3.7}$$

$$f_N(k) = f_N(n(k)) = \min\{\bar{\beta}_N\, v_N n_N(k), F_N\}, \tag{3.8}$$

$$n_0(k+1) = n_0(k) - f_0(k) + r_0(k). \tag{3.9}$$

In view of (3.1) a useful alternative to (3.7) is

$$f_i(k) = \min\{\bar{\beta}_i\, v_i n_i(k), F_i - w_{i+1}[n_{i+1}(k) - n^c_{i+1}], F_i\}, \tag{3.10}$$

and if cell $i+1$ is uncongested ($n_{i+1}(k) \le n^c_{i+1}$), (3.10) simplifies to

$$f_i(k) = \min\{\bar{\beta}_i\, v_i n_i(k), F_i\}. \tag{3.11}$$

Split ratios $\beta_1 \cdots, \beta_N$ are fixed. Assume stationary demands $r_i(k) \equiv r_i$. Each on-ramp *demand* vector $r = (r_0, \cdots, r_N)$ induces a unique equilibrium *flow* vector $f(r) = (f_0, \cdots, f_N)$ calculated as

$$f_0 = r_0, \tag{3.12}$$

$$f_i = \bar{\beta}_i\,(f_{i-1} + r_i), \quad 1 \le i \le N. \tag{3.13}$$

The function $r \mapsto f(r)$ defined by (3.12), (3.13) is bijective. We say that demand r is *feasible* if $0 \le f_i \le F_i$, $0 \le i \le N$; it is *strictly feasible* if $0 \le f_i < F_i$, $0 \le i \le N$. A strictly feasible demand induces an equilibrium flow with excess capacity in every cell.

State $n = (n_1, \cdots, n_N)$ is an *equilibrium* for a feasible demand r with induced flow $f = f(r)$, if the constant trajectory $n(k) \equiv n$ is a solution of (3.6)-(3.8):

$$f_i = \min\{\bar{\beta}_i \, v_i \, n_i, F_i - w_{i+1}(n_{i+1} - n^c_{i+1}), F_i\}, \quad 1 \le i \le N-1, \tag{3.14}$$

$$f_N = \min\{\bar{\beta}_N \, v_N \, n_N, F_N\}. \tag{3.15}$$

At equilibrium n, cell i is said to be *uncongested* if $0 \le n_i \le n^c_i$ and *congested* if $n_i > n^c_i$. The equilibrium n is *uncongested* if all cells are uncongested, otherwise it is congested.

Let $E = E(r)$ be the set of equilibria, i.e., all solutions of the system of equations (3.14)-(3.15), corresponding to a demand r. This section is devoted to characterizing $E(r)$ and the pattern of congested cells for each $n \in E(r)$. If r is not feasible, there is no solution to (3.14)-(3.15), so $E(r) = \emptyset$. Lemma 3.2.1 implies that $E(r) \neq \emptyset$ if r is feasible.

Lemma 3.2.1 *With feasible demand r, system (3.6)-(3.9) has a unique uncongested equilibrium $n^u(r)$.*

Proposition 3.2.1 *Suppose at equilibrium n, cell $i+1$ is uncongested and cells $i-j, \cdots, i$ are congested. Then*

$$f_i = F_i, \;\; \bar{\beta}_i^{-1} F_i - r_i = f_{i-1} < F_{i-1}, \cdots, f_{i-j-1} < F_{i-j-1}. \tag{3.16}$$

Corollary 3.2.1 *The same system with strictly feasible demand r has a unique equilibrium, so $E(r) = \{n^u\}$.*

The next result is a partial converse to Proposition 3.2.1.

Proposition 3.2.2 *Suppose $f_i = F_i, f_{i-1} < F_{i-1}, \cdots, f_{i-j} < F_{i-j}$. Suppose at equilibrium n, cell $i-j$ is congested. Then cells $i-j, \cdots, i-1, i$ are all congested at n.*

We say that i is a *bottleneck* cell for demand r (or induced flow f) if $f_i = F_i$. Suppose there are $K \geq 0$ bottleneck cells at $1 \leq I_1 < I_2 \cdots < I_K \leq N$. Partition the freeway into $1+K$ segments $S^1, \cdots, S^{K+1}$ comprising contiguous cells as follows:

$$S^1 = \{1, \cdots, I_1\}, \ \cdots, \ S^K = \{I_{K-1}+1, \cdots, I_K\}, \ S^{K+1} = \{I_K+1, \cdots, N\}. \tag{3.17}$$

If there are no bottleneck cells, $K = 0$, and $S^1 = \{1, \cdots, N\}$ is the entire freeway. On the other hand, if the most downstream cell is congested, $I_K = N$, and S^{K+1} is the empty segment.

Proposition 3.2.3 *The cells immediately downstream of the segments $S^1, \cdots, S^{K+1}$ are uncongested. Consequently, for $k = 1, \cdots, K$,*

$$f_{I_k} = \min\{\bar{\beta}_{I_k} v_{I_k} n_{I_k}, F_{I_k}\}. \tag{3.18}$$

Partition the N-dimensional state $n = (n_1, \cdots, n_N)$ into sub-vectors $n = (n^1, \cdots, n^{K+1})$ in conformity with the segments $S^1, \cdots, S^{K+1}$, so n^k has components $\{n_i, i \in S^k\}$. Since the equilibrium flow immediately upstream of segment S^k is known (it is equal to capacity) and

the cell immediately downstream of S^k is uncongested, the equilibrium conditions (3.14)-(3.15) partition into $1+K$ *decoupled* conditions, one for each segment. Thus,

n^k satisfies for $k = 1, \cdots, K$,

$$f_{I_k} = \min\{\bar{\beta}_{I_k} v_{I_k} n_{I_k}, F_{I_k}\}, \tag{3.19}$$

$$f_i = \min\{\bar{\beta}_i v_i n_i, F_i - w_{i+1}(n_{i+1} - n^c_{i+1}), F_i\}, \quad I_{k-1} < i \leq I_k; \tag{3.20}$$

n^{K+1} satisfies

$$f_N = \min\{\bar{\beta}_N \, v_N \, n_N, F_N\}, \tag{3.21}$$

$$f_i = \min\{\bar{\beta}_i \, v_i \, n_i, F_i - w_{i+1}(n_{i+1} - n^c_{i+1}), F_i\}, \quad I_K < i \leq N. \tag{3.22}$$

These decoupled conditions decompose the equilibrium set.

Proposition 3.2.4 *The set of equilibria $E(r)$ factors into the product set,*

$$E(r) = E^1(r) \times \cdots \times E^K(r) \times E^{K+1}(r), \tag{3.23}$$

in which $E^{K+1}(r)$ is the set of solutions n^{K+1} of (3.21)-(3.22) and $E^k(r)$ is the set of solutions n^k of (3.19)-(3.20) for $1 \leq k \leq K$.

We now fully characterize the components $E^1(r), \cdots, E^{K+1}(r)$. Recall that the flow in all non-bottleneck cells is strictly below capacity:

$$f_i < F_i, \quad i \notin \{I_1, \cdots, I_K\}. \tag{3.24}$$

Lemma 3.2.2 *$E^{K+1}(r) = \{n^{u,K+1}\}$ consists of a single point, the component of the uncongested equilibrium n^u corresponding to segment S^{K+1}. Hence $n^{u,K+1}$ is given by*

$$n_i^{u,K+1} = (\bar{\beta}_i v_i)^{-1} f_i, \quad I_K < i \leq N. \tag{3.25}$$

The next result gives an explicit expression for the equilibrium set $E^k(r)$, $1 \le k \le N$.

Lemma 3.2.3 *$n^k \in E^k(r)$ if and only if either (i) there is no congested segment at n^k and $n^k = n^{u,k}$, or (ii) there exists $j \in \{I_{k-1}+1, \cdots, I_k\}$ such that at n^k cells $I_{k-1}+1, \cdots, j-1$ are uncongested, cells $j, \cdots, I_k$ are congested and n^k is given by (see figure 3.3)*

$$n_i^k = n_i^u = (\bar{\beta}_i v_i)^{-1} f_i, \quad I_{k-1} < i < j+1 \tag{3.26}$$

$$n_i^k = n_i^c + w_i^{-1}(F_{i-1} - f_{i-1}), \quad j < i \le I_k \tag{3.27}$$

$$n_{j-1}^k \in [(\bar{\beta}_{j-1} v_{j-1})^{-1} f_{j-1}, n_{j-1}^c] \text{ and } n_j^k = n_j^c + w_j^{-1}(F_{j-1} - f_{j-1}), \text{ or} \tag{3.28}$$

$$n_{j-1}^k = n_{j-1}^u = (\bar{\beta}_{j-1} v_{j-1})^{-1} f_{j-1} \text{ and } n_j^k \in (n_j^c, n_j^c + w_j^{-1}(F_{j-1} - f_{j-1})] \tag{3.29}$$

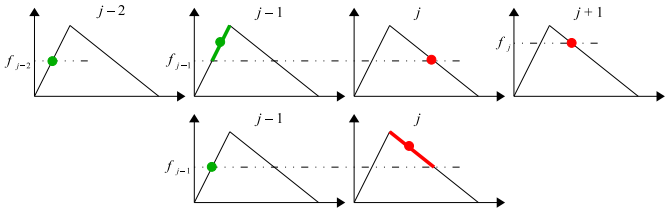


Figure 3.3: Equilibrium satisfying (3.26),(3.27) and (3.28) (top) or (3.29) (bottom).

Three densities appear in the expression of $E^k(r)$, namely $n_i^u = (\bar{\beta}_i v_i)^{-1} f_i$, the *uncongested* equilibrium density; n_i^c, the *critical* density; and the *congested* equilibrium density

$$n_i^{con} = n_i^c + w_i^{-1}(F_{i-1} - f_{i-1}). \tag{3.30}$$

By Lemma 3.2.3

$$E^k(r) = \{n^{u,k}\} \bigcup_{j \in S^k} E_j^k(r), \tag{3.31}$$

in which $E_j^k(r)$ is the set of n^k satisfying (3.26)-(3.29):

$$\begin{aligned} E_j^k(r) &= \{(n^u_{I_{k-1}+1}, \cdots, n^u_{j-1}, n_j, n^{con}_{j+1}, \cdots, n^{con}_{I_k}) \mid n_j \in (n^c_j, n^{con}_j]\} \\ &\bigcup \{(n^u_{I_{k-1}+1}, \cdots, n^u_{j-2}, n_{j-1}, n^{con}_j, \cdots, n^{con}_{I_k}) \mid n_{j-1} \in [n^u_{j-1}, n^c_{j-1}]\}. \end{aligned} \quad (3.32)$$

Observe that $f_{I_{k-1}+1} = F_{I_{k-1}+1}$, $n^u_{I_{k-1}+1} = (\bar{\beta}_{I_{k-1}+1} v_{I_{k-1}+1})^{-1} F_i = n^c_{I_{k-1}+1}$. So it follows from (3.32) that $n^{u,k} \in E^k_{I_k}(r)$. Hence (3.31) simplifies to

$$E^k(r) = \bigcup_{j \in S^k} E^k_j(r). \quad (3.33)$$

Observe next that the first set on the right hand side in (3.32) forms a straight line segment $E^k_{j,-}$ connecting the points

$$n^k(j-) = (n^u_{I_{k-1}+1}, \cdots, n^u_{j-1}, n^c_j, n^{con}_{j+1}, \cdots, n^{con}_{I_k}) \quad (3.34)$$

and

$$n^k(j) = (n^u_{I_{k-1}+1}, \cdots, n^u_{j-1}, n^{con}_j, \cdots, n^{con}_{I_k}). \quad (3.35)$$

Denote this line segment in terms of its endpoints as

$$E^k_{j-}(r) = (n^k(j-), n^k(j)]. \quad (3.36)$$

Similarly, the second set on the right hand side in (3.32) forms a straight line segment connecting the points $n^k(j)$ and

$$n^k(j+) = (n^u_{I_{k-1}+1}, \cdots, n^u_{j-2}, n^c_{j-1}, n^{con}_j, \cdots, n^{con}_{I_k}), \quad (3.37)$$

and denoted as

$$E^k_{j+}(r) = [n^k(j), n^k(j+)]. \quad (3.38)$$

The two line segments have exactly one point, $n^k(j)$, in common. Thus

$$E_j^k(r) = E_{j-}^k(r) \cup E_{j+}^k(r), \tag{3.39}$$

and, by comparing (3.34) and (3.37) one sees that

$$n^k(j+) = n^k(j-1-), \tag{3.40}$$

so that $E_j^k(r)$ and $E_{(j-1)}^k(r)$ have exactly this point in common. Lastly, since the densities $n_i^u \le n_i^c \le n_i^{con}$ are ordered, so are the endpoints:

$$\cdots \le n^k(j-) \le n^k(j) \le n^k(j+) = n^k((j-1)-) \le n^k(j-1) \le \cdots . \tag{3.41}$$

(For vectors x and y, $x \le y$ means $x_i \le y_i$ for all components i.)

Figure 3.4 depicts the projection of $E_j^k(r) = E_{j-}^k(r) \cup E_{j+}^k(r)$ on the two dimensional space spanned by n_{j-1}^k, n_j^k and the projection of $E_{j-1}^k = E_{(j-1)-}^k(r) \cup E_{(j-1)+}^k(r)$ on the space spanned by n_{j-2}^k, n_{j-1}^k. According to (3.40) the two highlighted points in the figure are the same.

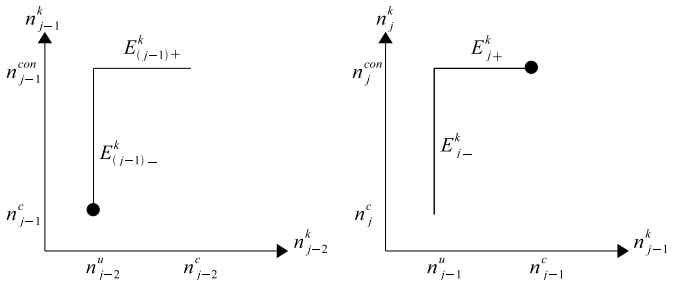


Figure 3.4: Projection of $E_{j-1}^k(r)$ on coordinates n_{j-2}^k, n_{j-1}^k (left) and projection of $E_j^k(r)$ on coordinates n_{j-1}^k, n_j^k (right).

Observe lastly that the straight line segments $E_{(j-1)-}$ and E_{j+} are aligned.

Theorem 3.2.1 follows from Proposition 3.2.4, and Lemmas 3.2.2 and 3.2.3.

Theorem 3.2.1 *Let r be a feasible demand, f the induced equilibrium flow, and $E(r)$ the equilibrium set. If r is strictly feasible, $E(r)$ consists of the unique uncongested equilibrium n^u. Otherwise, partition the freeway into segments $S^1, \cdots, S^{K+1}$ corresponding to the bottleneck cells $1 \leq I_1 < \cdots < I_K \leq N$. Then $E(r)$ is the direct product (3.23):*

$$E(r) = E^1(r) \times \cdots \times E^{K+1}(r).$$

Each $E^k(r)$ decomposes as the union (3.33):

$$E^k(r) = \bigcup_{j \in S^k} E^k_j(r), \ 1 \leq k \leq K \quad E^{K+1}(r) = \{n^{u,K+1}\}.$$

Each $E^k_j(r)$ is the union of two connected line segments, given by the 'closed form' expression (3.36), (3.38), (3.39). $E^k(r)$ is the union of $|S^k|$ connected straight line segments. ($|S_k| = |I_{k+1} - I_k|$ is the number of cells in S^k.)

Consecutive sets $E^k_j(r)$ and $E^k_{j+1}(r)$ have exactly one point in common, and they are ordered: if $n^k \in E^k_j(r)$ and $n'^k \in E^k_{j+1}$, then $n^k \geq n'^k$. In particular, the most congested equilibrium in $E^k(r)$ is $n^{con,k}$ with components $n^{con}_i, i \in S^k$, given by (3.30). Every $n^k \in E^k(r)$ lies between the uncongested equilibrium $n^{u,k}$ and $n^{con,k}$, i.e., $n^{u,k} \leq n^k \leq n^{con,k}$. Hence for all $n \in E(r)$,

$$n^u \leq n \leq n^{con},$$

in which the most congested equilibrium is $n^{con} = (n^{con,1}, \cdots, n^{con,K}, n^{u,K+1})$.

Lastly, $E(r)$ forms a connected, topologically closed surface of dimension K in the N-dimensional state space.

Proof. Only the last assertion needs proof, which follows from the observation that $E(r)$ is the product of $K+1$ sets, $E^1(r), \cdots, E^{K+1}(r)$, the last of which being a single point has

dimension 0, and each of the rest being a union of connected line segments has dimension 1. □

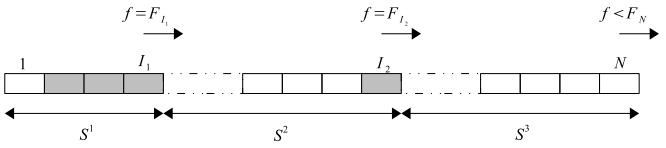


Figure 3.5: The demand induces two bottleneck cells and three segments. S^3 is uncongested. In the depicted equilibrium S^2 has one congested cell and S^1 has three congested cells.

Figure 3.5 illustrates the use of Theorem 3.2.1. The demand induces a flow that gives rise to bottlenecks at I_1, I_2 which partition the freeway into three segments S^1, S^2, S^3. S^3 is uncongested. An equilibrium determines the number of congested cells in the other segments. The figure illustrates an equilibrium in which one cell in S^2 and three cells in S^1 are congested (depicted by shaded rectangles); the others are uncongested. The congested cells must lie immediately upstream of the corresponding bottleneck.

3.3 Dynamic Behavior

Theorem 3.2.1 fully characterizes the equilibrium behavior of any CTM model. This section is devoted to the complete description of the qualitative behavior of all trajectories of the N-dimensional difference equation system (3.6)-(3.8). We assume a constant feasible demand r and write (3.6)-(3.8) as

$$n_i(k+1) = g_i(n(k)), \quad 1 \leq i \leq N. \tag{3.42}$$

Let $g = (g_1, \cdots, g_N)$. We will consider initial conditions

$$n \in \Sigma = \{n \mid 0 \leq n_i \leq \bar{n}_i, 1 \leq i \leq N\}. \tag{3.43}$$

Each initial condition $n(0) \in \Sigma$ generates a *trajectory* $\{n(k), k \geq 0\}$ according to $n(k+1) = g(n(k))$.

For two vectors x, y in R^N, write

$$\begin{aligned} x \leq y &\Leftrightarrow x_i \leq y_i, \\ x < y &\Leftrightarrow x \leq y, x \neq y, \\ x \ll y &\Leftrightarrow x_i < y_i. \end{aligned}$$

Following [57] say that g is *strictly monotone* if, for $x, y \in \Sigma$,

$$x < y \Rightarrow g(x) < g(y);$$

g is *strongly monotone* if

$$x \leq y \Rightarrow g(x) \ll g(y).$$

Lemma 3.3.1 *The map g is strictly monotone, but it is not strongly monotone.*

[57] surveys the theory of monotone maps. The most powerful results, however, require strong monotonicity, and do not apply to CTM.

Let the equilibrium flow induced by the demand r result in bottlenecks at $1 \leq I_1 < \cdots < I_K \leq N$, and let $S^1, \cdots, S^{K+1}$ be the corresponding freeway partition. By Theorem 3.2.1 every equilbrium lies between the uncongested equilibrium n^u and the most congested equilibrium n^{con},

$$n^u \leq n \leq n^{con}, \quad n \in E(r). \tag{3.44}$$

Let $\hat{n}(k), k \geq 0$ be the trajectory starting with the empty freeway, $\hat{n}(0) = 0$, and let $\bar{n}(k), k \geq 0$ be the trajectory starting with the completely jammed freeway, $\bar{n}_i(0) = \bar{n}_i, 1 \leq i \leq N$. Let $n(k)$ be a trajectory starting in any state $n(0) \in \Sigma$. The next result shows how much monotonicity of g constrains the trajectories of the CTM model.

Lemma 3.3.2 *(i) Every trajectory lies between $\{\hat{n}(k)\}$ and $\{\bar{n}(k)\}$:*

$$\hat{n}(k) \leq n(k) \leq \bar{n}(k), \quad k \geq 0. \tag{3.45}$$

(ii) The trajectory starting with the empty freeway converges to the uncongested equilibrium n^u:

$$\lim_{k\to\infty} \hat{n}(k) = n^u. \tag{3.46}$$

(iii) The trajectory starting with the completely jammed freeway converges to the most congested equilibrium n^{con}:

$$\lim_{k\to\infty} \bar{n}(k) = n^{con}. \tag{3.47}$$

Lemma 3.3.2 leads to Theorem 3.3.1: If the demand is strictly feasible, then n^u is a globally, asymptotically stable equilibrium.

Theorem 3.3.1 *Suppose r is strictly feasible. Then every trajectory converges to n^u.*

Proof. By Lemma 3.2.1 $E(r) = \{n^u\}$, so $n^{con} = n^u$. Hence both $\hat{n}(k)$ and $\bar{n}(k)$ converge to n^u. By (3.45), every trajectory $n(k)$ converges to n^u as well. □

If r is not strictly feasible, the equilibrium set $E(r)$ is infinite and there is no easy way to analyze how trajectories behave. The main result of this section, Theorem 3.3.3, is that

every trajectory converges to some equilibrium. Before getting into the complexities of the proof, we study three examples.

Examples

Example 1 is a freeway with two identical cells, each one mile long. The fundamental diagram, equilibrium flow, and equilibrium set E are shown in Figure 3.6. The critical density $n_i^c = 100$ veh/mile; the jam density $\bar{n}_i = 400$ veh/mile; free flow speed $v = 60$ mph and the congestion wave speed $w = 20$ mph. The demand vector $r = (r_0 = 4800, r_1 = 0, r_2 = 1200)$, all in vehicles per hour (vph). The upstream flow $r_0 = f_0 = 4800$ vph, and $f_2 = F_2 = 6000$ vph. Thus cell 2 is the only bottleneck. The uncongested equilibrium $n^u = (80, 100)$ and the most congested equilibrium $n^{con} = (160, 160)$. By Theorem 3.2.1, the equilibrium set E consists of two straight line segments shown in the figure (also see Figure 3.4).

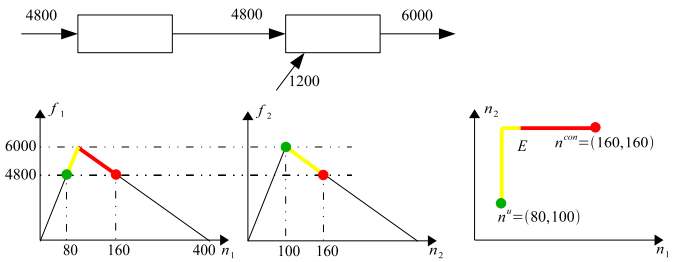


Figure 3.6: Freeway, equilibrium flows, fundamental diagram, and equilibrium set E of Example 1.

The phase portrait of Figure 3.7 displays the orbits of the two-dimensional state with initial conditions on the boundary of the square $\Sigma = [0, 400] \times [0, 400]$. (An orbit is the set of

states $\{n(k) \mid k \geq 0\}$ traversed by a trajectory $k \to n(k)$.) We analyze the orbit structure displayed in Figure 3.7. The observations made below hold in general.

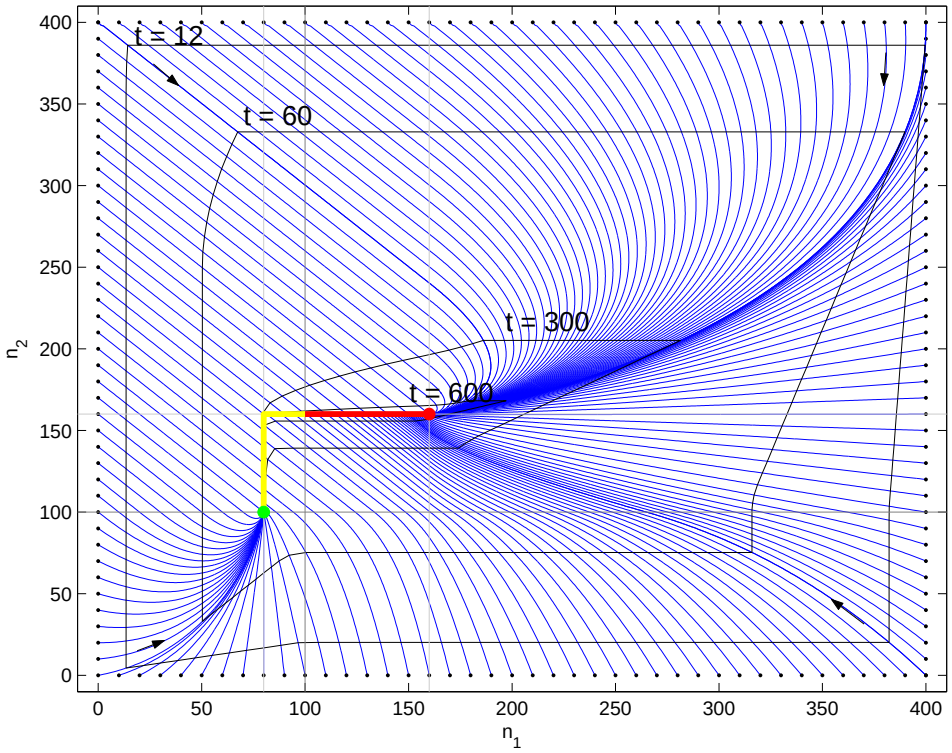


Figure 3.7: Equilibrium set and orbits of Example 1.

1. Every trajectory converges to an equilibrium point in E. As a consequence, the state space Σ is partitioned as

$$\Sigma = \bigcup_{n \in E} \Sigma(n),$$

in which $\Sigma(n)$ is the set of all initial states whose trajectories converge to the equilibrium n. By monotonicity $\Sigma(n^u)$ includes all initial states $n \leq n^u$, and $\Sigma(n^{con})$ includes all initial states $n \geq n^{con}$. By contrast, for all other equilibrium states $\Sigma(n)$ is simply a one-dimensional manifold. (In the general case, $\Sigma(n)$ is a $(N-K)$-dimensional

manifold, Corollary 3.3.1.)

2. The figure shows four equi-time contour plots, labeled $k = 12s, \cdots, 600s$. For example, the contour plot $k = 60s$ is the set of points reached by all trajectories at $k = 60$ seconds. As k increases, the contour plots converge towards the equilibrium set E. As might be expected, the contours initially converge rapidly and the convergence slows down as E is approached. More interestingly, consider the orbit going through the state $n = (50, 340)$ on the $k = 60$ contour. In this state cell 2 is congested but cell 1 is not. However, by time 200 (whose contour plot is not shown) the state has moved to approximately $(150, 250)$, indicating both cell are congested. The time difference of $200 - 60 = 140$ seconds is roughly predictable: because the congestion wave speed is 20 mph it takes about 3 minutes for the congestion wave to travel the one mile-long cell.

3. According to Theorem 3.2.1 the equilibrium set is ordered: if n, n' are two equilibria, either $n \leq n'$ or $n' \leq n$. Consequently, downstream cells must get congested before an upstream cell. As seen in the figure, every trajectory in which cell 1 is getting congested also congests cell 2.

4. All equilibria support the same equilibrium flows. However at equilibrium n^u the speed is $v = 60$ mph throughout, whereas in n^{con} the speed is 4800/160 (flow/density) or 30 mph. Thus, although both n^u and n^{con} achieve the same throughput, the freeway travel time in n^u is one-half of that in n^{con}.

In Example 2 the flow r_0 is slightly reduced from 4800 to 4750 vph, so the demand becomes

strictly feasible and the equilibrium set collapses to the single uncongested equilibrium n^u. The resulting phase portrait in Figure 3.8 can be compared with Figure 3.7. The trajectories are nearly identical, except that when they approach the equilibrium set of Example 1 they turn and converge to $n^u \approx (80, 100)$.

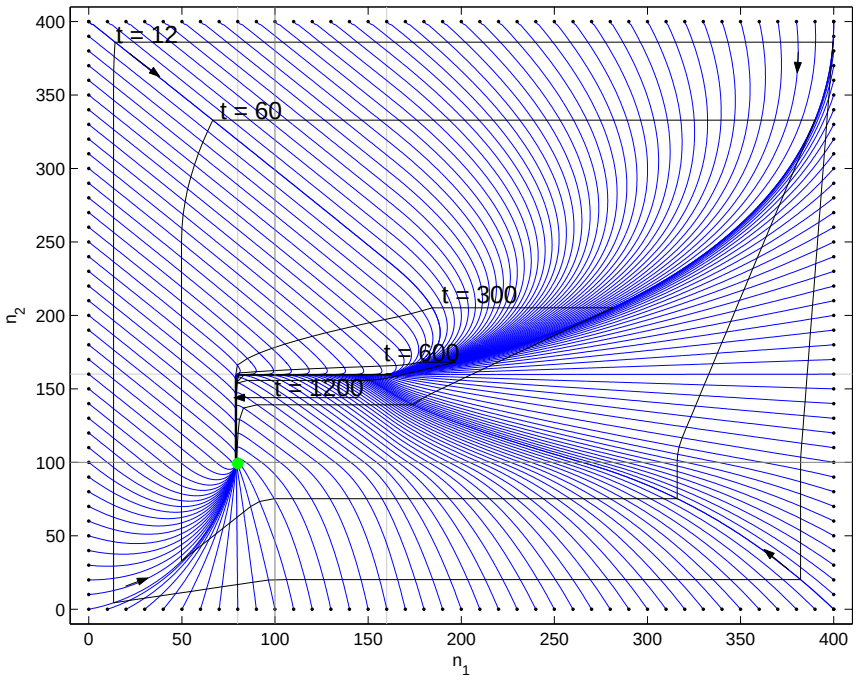


Figure 3.8: Equilibrium set and orbits of Example 2.

Example 3 shown in Figure 3.9 is a modification of Example 1 in that there are three identical cells. The fundamental diagram is the same as in Example 1. The demand $r_0 = 4800, r_1 = r_2 = 0, r_3 = 1200$. Again the most downstream cell, cell 3, is the only bottleneck. The equilibrium set now comprises three straight line segments, connecting the uncongested equilibrium $n^u = (80, 80, 100)$ and the most congested equilibrium $n^{con} = (160, 160, 160)$. The orbit structure supports the observations made earlier: although it is less apparent in the figure, $\Sigma(n)$ is a 2-dimensional manifold if $n \neq n^u, n^{con}$.

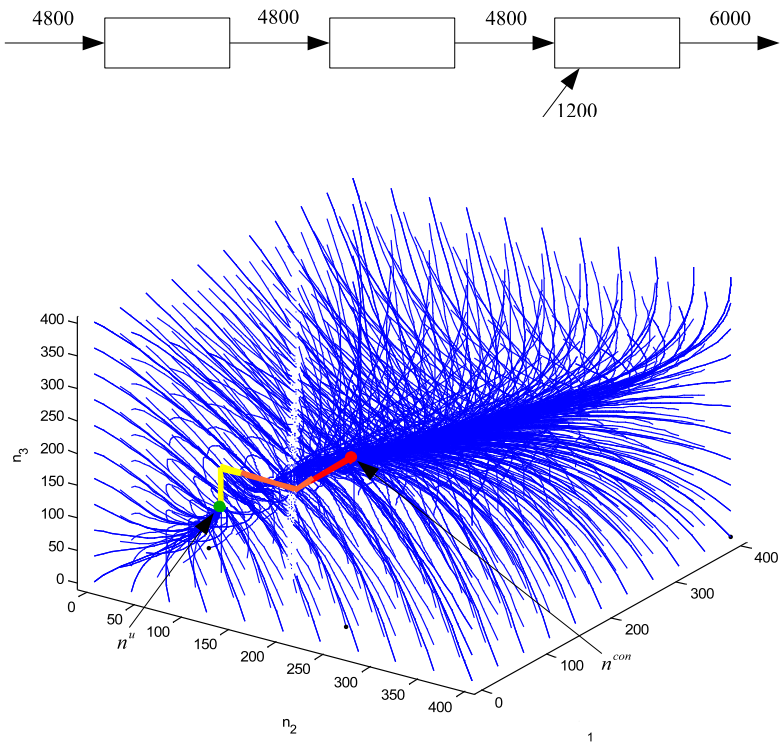


Figure 3.9: Freeway, equilibrium set and orbits of Example 3.

We resume the general discussion. As before let r be a demand vector and ϕ the resulting equilibrium flow vector, i.e., (see (3.12), (3.13))

$$\phi_0 = r_0, \quad \phi_i = \bar{\beta}_i(\phi_{i-1} + r_i), \quad 1 \leq i \leq N. \tag{3.48}$$

Let $1 \leq I_1 < \cdots < I_K \leq N$ be the bottlenecks, and $S^1, \cdots, S^{K+1}$ the corresponding freeway partition. By Theorem 3.2.1 the equilibrium set decomposes as

$$E(r) = E^1(r) \times \cdots \times E^K(r) \times \{n^{u,K+1}\}. \tag{3.49}$$

Let $\hat{n}(k), k \geq 0$, be the trajectory starting at 0 and converging to n^u. Let $\bar{n}(k)$ be the trajectory starting at $\bar{n}$ and converging to n^{con}.

Fix an initial condition $n \in \Sigma$ and let $n(k), k \geq 0$, be the trajectory starting at n.

We recall some facts from the general theory of dynamical systems. The *ω-limit* set of n is the set of all limit points of the trajectory $\{n(k)\}$:

$$\omega(n) = \{p \in \Sigma \mid \text{there is a subsequence } k_m \text{ with } \lim_{m\to\infty} n(k_m) = p\}.$$

$\omega(n)$ is non-empty, compact, and *invariant*, i.e., if $p \in \omega(n)$ the trajectory starting at p stays within $\omega(n)$. Furthermore the trajectory converges to $\omega(n)$, i.e., $\lim_k d(n(k), \omega(n)) = 0$, with $d(x, \omega(n)) = \min\{|x - p| \mid p \in \omega(n)\}$.

Our objective is to prove that the trajectory $\{n(k)\}$ converges to an equilibrium, which is achieved in two steps. The first step shows that $\omega(n)$ always contain an equilibrium (Lemma 3.3.4). The second step shows that every equilibrium is stable (Theorem 3.3.2).

We adopt the following notation. For any $p \in \Sigma$,

$$f_i(p) = \begin{cases} \min\{\bar{\beta}_i v_i p_i, w_{i+1}[\bar{n}_{i+1} - p_{i+1}], F_i\}, & 1 \leq i < N \\ \min\{\bar{\beta}_N v_N p_N, F_N\}, & i = N \end{cases},$$

and

$$f_i(k) = f_i(n(k)).$$

Lemma 3.3.3 *(i) Suppose $n^u \leq p \leq n^{con}$. Then*

$$f_i(p) \geq \phi_i, \text{ all } i, \text{ and } f_i(p) = F_i, \ i \in \{I_1, \cdots, I_K\}. \tag{3.50}$$

(ii) If $p \in \omega(n)$, $n^u \leq p \leq n^{con}$.

(iii) Along the trajectory $\{n(k)\}$

$$\liminf_{k\to\infty} f_i(k) \geq \phi_i, \text{ all } i, \text{ and } \lim_{k\to\infty} f_i(k) = F_i, \ i \in \{I_1, \cdots, I_K\}. \tag{3.51}$$

To simplify the discussion we assume that n_0, the upstream ramp queue, is always so large as to maintain

$$f_0(k) \equiv r_0 = \phi_0, \quad k \geq 0. \tag{3.52}$$

Lemma 3.3.4 $\omega(n) \cap E(r) \neq \emptyset$.

Recall the definition of (Lyapunov) stability: An equilibrium n^e is *stable* if for every $\epsilon > 0$ there is $\delta > 0$ such that $|n - n^e| < \delta$ implies $|n(k) - n^e| < \epsilon$ for all k, in which $\{n(k)\}$ is the trajectory starting at n.

Fix an equilibrium n^e. By Theorem 3.2.1 n^e has the form

$$n^e = (n^{e,1}, \cdots, n^{e,K}, n^{e,K+1}),$$

with $n^{e,K+1} = n^{u,K+1}$, $n^{e,m} \in E_j^m(r)$ for some $j \in S^m$, $1 \leq m \leq K+1$.

Lemmas 3.3.5-3.3.6 will prove that if $|n - n^e| < \delta$, and $n(k) = (n^1(k), \cdots, n^K(k), n^{K+1}(k))$, $k \geq 0$ is the trajectory starting at n, then there exists an equilibrium $\tilde{n}^e$, possibly different from n^e, such that

$$|\tilde{n}^{e,m} - n^{e,m}| < \epsilon, \quad \lim_{k\to\infty} n^m(k) = \tilde{n}^{e,m}, \quad 1 \leq m \leq K+1, \tag{3.53}$$

Lemma 3.3.5 *(3.53) holds for $m = K+1$. In fact*

$$\lim_{k\to\infty} n^{K+1}(k) = n^{u,K+1} = n^{e,K+1}. \tag{3.54}$$

Lemma 3.3.6 *Suppose (3.53) holds for $m - 1 \geq 0$. Then it holds for m.*

Theorem 3.3.2 *Every equilibrium $n^e \in E(r)$ is stable. In fact for $\epsilon > 0$ there is $\delta > 0$ such that if $|n - n^e| < \delta$, the trajectory $\{n(k)\}$ starting at n converges to an equilibrium $\tilde{n}^e$ with $|\tilde{n}^e - n^e| < \epsilon$.*

Proof Lemmas 3.3.5-3.3.6 prove the second part of the assertion which implies stability. □

Figure 3.7 illustrates Theorem 3.3.1. Trajectories starting close to an equilibrium all converge to some nearby equilibrium.

Theorem 3.3.3 *The CTM model is a convergent system, i.e. every trajectory converges to some equilibrium in $E(r)$.*

Proof Consider any trajectory $\{n(k)\}$. By Lemma 3.3.4 there is an equilibrium n^e and a subsequence $\{k_m\}$ along which $n(k_m) \to n^e$ as $m \to \infty$. By Theorem 3.3.2 the trajectory must converge to this equilibrium. □

Recall that the stable manifold $\Sigma(n^e)$ of an equilibrium $n^e \in E(r)$ comprises all $n \in \Sigma$ whose trajectories converge to n^e. The next result characterizes the orbit structure.

Corollary 3.3.1 *If r is strictly feasible, $E(r) = \{n^u\}$ and $\Sigma(n^u) = \Sigma$. If r is not strictly feasible, $E(r)$ is a K-dimensional manifold and $\Sigma(n^e)$ is a $(N-K)$-dimensional manifold for $n^e \neq n^u, n^{con}$, whereas $\Sigma(n^u), \Sigma(n^{con})$ are N-dimensional manifolds with boundary.*

3.4 Implications for Ramp Metering

We explore two implications for ramp metering. The first considers the case when the demand vector r is *infeasible*, i.e., the associated equilibrium flow ϕ given by (3.48) is such that it exceeds the capacity in some cell. Peak hour demand may be infeasible in this sense. We begin with an example to illustrate the issues.

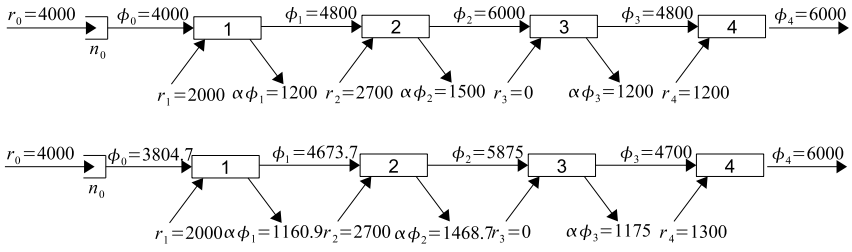


Figure 3.10: Freeway, on-ramp and off-ramp flows of Example: feasible demand (top); excess demand (bottom).

Example. The upper part of Figure 3.10 displays a freeway with four identical cells, each with capacity 6000 vph. The demand vector $r = (r_0 = 4000, r_1 = 2000, r_2 = 2700, r_3 = 0, r_4 = 1200)$. All split ratios are the same: $\beta_i = \beta = 0.2$, so $\bar{\beta} = 0.8$ and $\alpha = \beta[\bar{\beta}]^{-1} = 0.25$. The demand r is feasible and the equilibrium flow $\phi = (\phi_0 = 4000, \phi_1 = 4800, \phi_2 = 6000, \phi_3 = 4800, \phi_4 = 6000)$. The off-ramp flow in cell i is $\alpha\phi_i$. Cells 2 and 4 are bottleneck cells, with equilibirum flows equal to capacity.

Now consider the demand $\tilde{r}$ in which $\tilde{r}_4 = 1300 > r_4$ and $\tilde{r}_i = r_i, 0 \le i \le 3$. This demand is not feasible because it would increase ϕ_4 to $\phi_3 + \tilde{r}_4 = 6100$, which exceeds capacity. Evidently, the increased on-ramp flow in cell 4 will create congestion in cell 4 and force a reduction in the flow out of cell 3 from $\phi_3 = 4800$ to $\tilde{\phi}_3 = 4700$. This reduction from ϕ_3 to

$\tilde{\phi}_3$ is achieved by a reduction in the flow from cell 2 from $\phi_2 = 6000$ to $\tilde{\phi}_2 = 5875$, which in turn reduces the flow from cell 1 from $\phi_1 = 4800$ to $\tilde{\phi}_1 = 4673.75$, and ultimately the flow from cell 0 from $\phi_0 = 4000$ to $\tilde{\phi}_0 = 3804.7$. As a result the on-ramp queue n_0 will grow at the rate of $4000 - 3804.6875 = 195.3125$ vph. All cells will become congested.

The reductions in the equilibrium flow from ϕ to $\tilde{\phi}$ will proportionately reduce the discharge at the off-ramps from $\alpha\phi_i$ to $\alpha\tilde{\phi}_i$. The new equilibrium flows are displayed in the lower part of the figure.

The example suggests some observations.

1. The infeasible demand $\tilde{r}$ leads to a unique equilibrium flow $\tilde{\phi}$. This is the flow corresponding to the feasible demand $\tilde{r}^f$, which is the same as $\tilde{r}$, except that the upstream flow is reduced from $\phi_0 = 4000$ to $\tilde{\phi}_0 \approx 3804.7$. The system converges to the (unique) most congested equilibrium corresponding to $\tilde{r}^f$.

2. The reduction in the flow at the upstream ramp of about $196 = 4000 - 3804$ vph is *more* than the 'excess' demand of $1300 - 1200 = 100$ vph at the ramp in cell 4. Suppose that we meter the on-ramp in cell 4 and admit only 1200 vph. The queue at this ramp will now grow at 100 vph, but the resulting equilibrium flow and the off-ramp discharges will be the same as in the top of the figure; hence the total discharge will be *higher* by $196 - 100 = 96$ vph.

3. Figure 3.11 shows the phase portrait of the freeway considered in Figure 3.6. The figure also displays the equilibrium set $E(\tilde{r}^f)$. All of the trajectories converge to the most congested

equilibrium in $E(\tilde{r}^f)$. There is a pleasing symmetry with the case of strictly feasible demand, in which every trajectory converges to the uncongested equilibrium as in Figure 3.8.

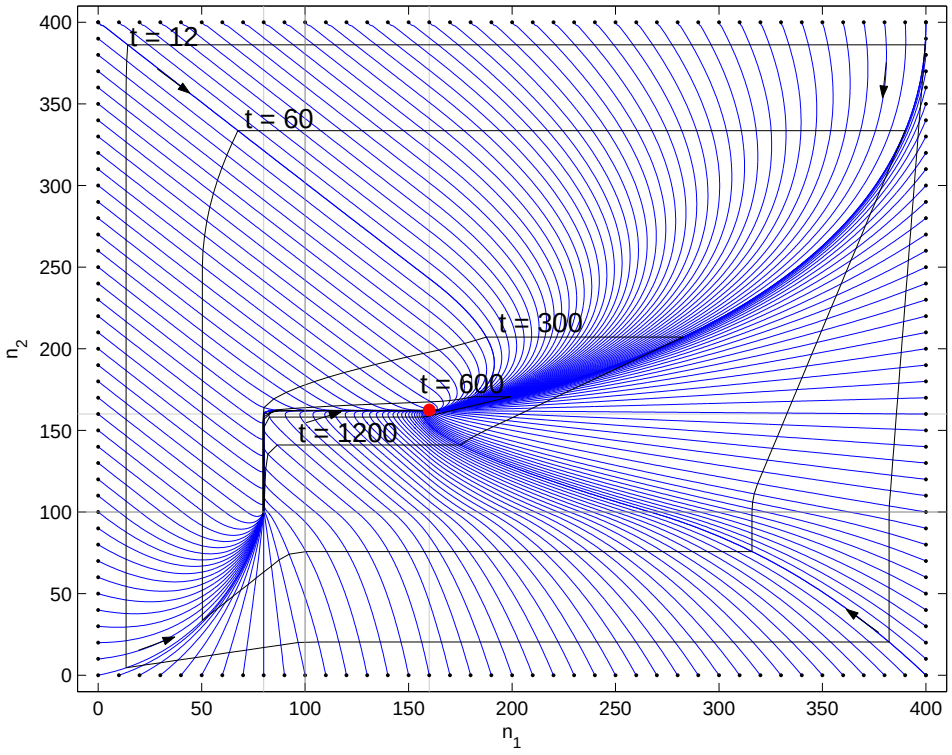


Figure 3.11: Orbits of the infeasible demand example.

The next result places the example above in a general setting. The freeway structure is the same as in sections 3.1-3.3. Let $r = (r_0, \cdots, r_N)$ be a demand vector. Let ϕ be the solution of (3.48):

$$\phi_0 = r_0, \quad \phi_i = \bar{\beta}_i(\phi_{i-1} + r_i), \quad 1 \leq i \leq N.$$

Suppose that r is *infeasible*, so that $\phi_i > F_i$ for some i.

To simplify the notation we make two assumptions. First $\phi_N > F_N$, and if $r_N = 0$ the demand becomes feasible. Second, if $r_0 = 0$ (zero inflow from the upstream ramp) the

demand again becomes feasible.

Since $\phi_N > F_N$, under demand r the entire freeway will become congested as in the example. Since with $r_0 = 0$ the demand is feasible,

$$\tilde{r}_0 = \max\{\rho \geq 0 \mid \text{ the demand } (\rho, r_1, \cdots, r_N) \text{ is feasible}\} \tag{3.55}$$

is well-defined, i.e., $\tilde{r}_0 \geq 0$. Since with $r_N = 0$ the demand is feasible,

$$\hat{r}_N = \max\{\rho \geq 0 \mid \text{ the demand } (r_0, \cdots, r_{N-1}, \rho) \text{ is feasible}\} \tag{3.56}$$

is similarly well-defined.

Theorem 3.4.1 *(i) $\tilde{r}_0 < r_0$ is the largest upstream flow for which the demand $\tilde{r} = (\tilde{r}_0, \cdots, r_{N-1}, r_N)$ is feasible. The corresponding equilibrium flow $\tilde{\phi}$ is*

$$\tilde{\phi}_0 = \tilde{r}_0, \quad \phi_i = \bar{\beta}_i(\tilde{\phi}_{i-1} + r_i), \quad 1 \leq i \leq N.$$

(ii) With demand r, under the no-metering strategy every trajectory converges to the (unique) most congested equilibrium $n^{con} \in E(\tilde{r})$ corresponding to demand $\tilde{r}$. Moreover, the queue $n_0(k)$ at the upstream ramp grows indefinitely at the rate of $(r_0 - \tilde{r}_0)$ vehicles per period.
(iii) $\hat{r}_N < r_N$ is the largest flow for which the demand $\hat{r} = (r_0, r_1, \cdots, \hat{r}_N)$ is feasible. The corresponding equilibrium flow $\hat{\phi}$ is

$$\hat{\phi}_0 = r_0, \ \hat{\phi}_i = \bar{\beta}_i(\hat{\phi}_{i-1} + r_i), \ 1 \leq i \leq N-1, \ \hat{\phi}_N = \bar{\beta}_N(\hat{\phi}_{N-1} + \hat{r}_N).$$

Under the ramp metering strategy that reduces the on-ramp flow in cell N from r_N to $\hat{r}_N$, every trajectory converges to some equilibrium in $E(\hat{r})$. The queue at the on-ramp in cell N

grows indefinitely at the rate of $(r_N - \hat{r}_N)$ *vehicles per period.*

(iv) Flows under the ramp-metering strategy are larger throughout the freeway:

$$\tilde{\phi}_i < \hat{\phi}_i,\ 1 \le i < N \text{ and } \tilde{\phi}_N = \hat{\phi}_N = F_N.$$

Suppose $\beta_i > 0$ *for some* $1 \le i < N$*, so that there is non-zero off-ramp flow in at least one cell. Then the total discharge under the ramp-metering strategy is strictly larger than under the no-metering strategy. Moreover,*

$$\mu = \frac{r_0 - \tilde{r}_0}{r_N - \hat{r}_N} = (\bar{\beta}_1 \cdots \bar{\beta}_{N-1})^{-1} > 1. \tag{3.57}$$

Proof. (i) follows from (3.55) and (3.48). Since the entire freeway becomes congested under r, every trajectory converges to $n^{con}(\tilde{r})$ by (3.47) and, by (i), vehicles accumulate at the upstream ramp at the rate of $(r_0 - \tilde{r}_0)$ per period. This proves (ii).

To prove (iii) we solve (3.48) recursively for $\tilde{r}$ and $\hat{r}$, setting $\bar{\beta}_0 = 1$, to get

$$\tilde{\phi}_i = \sum_{j=1}^{i} (\bar{\beta}_i \cdots \bar{\beta}_j) r_j + (\bar{\beta}_0 \cdots \bar{\beta}_i)\tilde{r}_0,\ 1 \le i \le N, \tag{3.58}$$

$$\hat{\phi}_i = \begin{cases} \sum_{j=1}^{i} (\bar{\beta}_i \cdots \bar{\beta}_j) r_j + (\bar{\beta}_0 \cdots \bar{\beta}_i) r_0,\ 1 \le i \le N-1 \\ \bar{\beta}_N \hat{r}_N + \sum_{j=1}^{N-1} (\bar{\beta}_j \cdots \bar{\beta}_N) r_j + (\bar{\beta}_0 \cdots \bar{\beta}_N) r_0,\ i = N \end{cases}. \tag{3.59}$$

Since $\tilde{r}_0 < r_0$ it follows from (3.58)-(3.59) that $\tilde{\phi}_i < \hat{\phi}_i,\ 0 \le i < N$. Also, since $\hat{r}_N$ is the largest flow that keeps $\hat{\phi}_N \le F_N$, it must be that $\hat{\phi}_N = F_N$. Similarly $\tilde{\phi}_N = F_N$. Hence if $\beta_i > 0$ for some $1 \le i < N$, then $\beta_i \hat{\phi}_i > \beta_i \tilde{\phi}_i$, i.e., the off-ramp discharge under

ramp-metering is strictly larger in at least one cell. Lastly, from (3.58) and (3.59),

$$\begin{aligned} F_N = \tilde{\phi}_N &= \sum_{j=1}^{N} (\bar{\beta}_j \cdots \bar{\beta}_N) r_j + (\bar{\beta}_0 \cdots \bar{\beta}_N) \tilde{r}_0, \\ F_N = \hat{\phi}_N &= \bar{\beta}_N \hat{r}_N + \sum_{j=1}^{N-1} (\bar{\beta}_j \cdots \bar{\beta}_N) r_j + (\bar{\beta}_0 \cdots \bar{\beta}_N) r_0, \end{aligned}$$

which, upon subtraction, gives

$$\bar{\beta}_N (r_N - \hat{r}_N) = (\bar{\beta}_0 \cdots \bar{\beta}_N)(r_0 - \tilde{r}_0),$$

and so

$$r_N - \hat{r}_N = (\bar{\beta}_1 \cdots \bar{\beta}_{N-1})(r_0 - \tilde{r}_0),$$

which implies (3.57) because $\bar{\beta}_i < 1$ for at least one i. □

Theorem 3.4.1 prompts some observations.

1. The discussion of infeasible demand above assumes that the on-ramp flow in a cell takes priority over the flow from the upstream cell: the latter cannot block an on-ramp flow, even if the cell is congested. This priority is implicit in the treatment of $r_i(k)$ in (3.2).

2. The unserved demand under the ramp metering strategy is $(r_N - \hat{r}_N)$ vehicles per period; the unserved demand under the no-metering strategy is $(r_0 - \tilde{r}_0)$. By (3.57), the no-metering strategy *magnifies* the unserved demand under the ramp strategy by $\mu = (\bar{\beta}_1 \cdots \bar{\beta}_{N-1})^{-1}$. The larger are the split ratios, the larger is the 'multiplier' μ, and worse is the no-metering strategy. (In the example of Figure 3.10 $\mu = (0.8 \times 0.8 \times 0.8)^{-1} \approx 2$.)

3. The ramp metering strategy increases speed in every cell i (hence reduces travel time). Because $\hat{\phi}_i > \tilde{\phi}_i$ and the freeway is congested under the no-metering strategy, the funda-

mental diagram implies that the density $\hat{n}_i < \tilde{n}_i$ which, in turn, implies that the speed (= flow/density) under ramp metering is higher: $\hat{\phi}_i/\hat{n}_i > \tilde{\phi}_i/\tilde{n}_i$.

4. Because of (3.57) the total travel time under the no-metering strategy grows arbitrarily larger than under the no-metering strategy.

5. An intuitive explanation of the increased off-ramp discharge under the ramp metering strategy might be that the no-metering strategy creates a congestion "queue" that blocks the off-ramps. This explanation is too crude. Note that under the ramp-metering strategy, the system can converge to any equilibrium in $E(\hat{r})$, including the most congested equilibrium $n^{con}(\hat{r})$, and under the no-metering strategy it converges to the most congested equilibrium $n^{con}(\tilde{r})$. Thus the entire freeway may be congested under both strategies. Nevertheless, the flows in every cell, and hence the off-ramp flows, are larger under the ramp metering strategy. Thus a more accurate (but less intuitive) explanation is that the congestion queue under ramp metering "moves faster" than the queue under the no-metering strategy.

While Theorem 3.4.1 is intuitively evident, the second implication of the theory *is* surprising: Theorem 3.4.2 says that ramp metering can reduce total travel time even when the demand is feasible.

Fix a feasible (but not strictly feasible demand) r; let ϕ be its equilibrium flow given by (3.48) and $E(r)$ its equilibrium set. Recall that $f(n^e) = \phi$ for all $n^e \in E(r)$.

To simplify the notation we assume that under r the only bottleneck is cell N; hence $\phi_N = F_N, \phi_i < F_i,\ 1 \le i < N,\ \phi_0 = r_0 \le F_0$. Suppose the freeway is initially in a congested equilibrium $n(0) = n^e$ in which cells $j, \cdots, N$ are congested for some $1 \le j < N$, with

$n_i^e(0) = n_i^{con}, j \le i \le N$, and cells $1, \cdots, j-1$ are uncongested with $n_i^e = n_i^u, 1 \le i < j$. For any $p \in \Sigma$ write $n^u \prec p \prec n^e$ if

$$n_i^u < p_i < n_i^e,\ j \le i \le N, \text{ and } n_i^u = p_i = n_i^e,\ 1 \le i < j. \tag{3.60}$$

Lemma 3.4.1 refines Lemma 3.3.3(i).

Lemma 3.4.1 *If* $n^u \prec p \prec n^e$,

$$f_N(p) = \phi_N = F_N, \quad f_i(p) > \phi_i,\ j \le i < j, \text{ and } f_i(p) = \phi_i,\ i > j. \tag{3.61}$$

We assume strictly positive demand in the congested cells, so

$$\rho = \min\{r_i,\ j \le i \le N\} > 0.$$

We construct a ramp metering strategy that selects the on-ramp flow values as follows:

$$r_i(k) = \begin{cases} r_i - \mu_i(k), & j \le i \le N \\ r_i, & 1 \le i < j \end{cases}. \tag{3.62}$$

(The $\{\mu_i\}$ are specified below in (3.87).) Denote by $p(k), k \ge 0$, the trajectory starting at $p(0) = n^e$ under the metering strategy (3.62).

Lemma 3.4.2 *There is a finite time horizon* K *and a metering strategy* $\{\mu_i(k),\ k = 0, \cdots, K\}$ *such that the resulting (controlled) trajectory* $p(k),\ k = 0, \cdots, K$, *satisfies*

$$n^u \prec p(k) \prec n^e,\ k = 1, \cdots, K-1, \tag{3.63}$$

and

$$p(K) = n^u. \tag{3.64}$$

In particular, the ramp metering strategy steers the state from the initial congested equilibrium n^e to the uncongested equilibrium n^u.

Theorem 3.4.2 *Suppose the freeway begins in a congested equilibrium n^e in which cells $1, \cdots, j-1$ are uncongested and cells $j, \cdots, N$ are congested. Then there exists a ramp metering strategy over a finite horizon K at the end of which the freeway is in the uncongested equilibrium n^u. Furthermore, the flows during $k = 0, \cdots, K$ are larger than the equilibrium flows. Finally, if the split ratio $\beta_i > 0$ for some $j \leq i \leq N$, then the total discharge flow is strictly larger and the total travel time is strictly smaller than in the no-metering strategy.*

Proof. By Lemma 3.4.2 in each cell i the flow $f_i(p(k)) > \phi_i$ for at least one k. Hence the difference in the total discharge

$$\sum_{k=0}^{K} \beta_i \bar{\beta}_i^{-1} [f_i(p) - \phi_i] > 0,$$

from which the assertion follows. □

Two observations are worth making.

1. If the split-ratios in the congested cells are all zero, $\beta_i = 0,\ i = j, \cdots, N$, the ramp metering strategy does not increase the total discharge, but it moves the system to the uncongested equilibrium n^u by 'moving' the 'excess' vehicles $\sum_{i=j}^{N} [n_i^e - n_i^u]$ from the congested cells to their on-ramps. The resulting total travel time is unchanged but traffic in the freeway moves at free flow speeds. If some of the traffic in the queues is diverted to alternative routes, perhaps along arterials, there will be a decline in total travel time just as with non-zero split-ratios.

2. There is another compelling reason for maintaining the freeway in free flow. The example of Figure 3.6 illustrates a common situation in which the congestion density of 160 vehicles/mile (and speed of 30 mph) compares with the uncongested density of 80 vehicles/mile (and speed of 60 mph) for a three-lane freeway. Storing the 80 additional vehicles would require a 3/4 mile-long one-lane on-ramp (at 50 feet vehicle spacing). Clearly congestion causes the freeway to be used as a very expensive parking place.

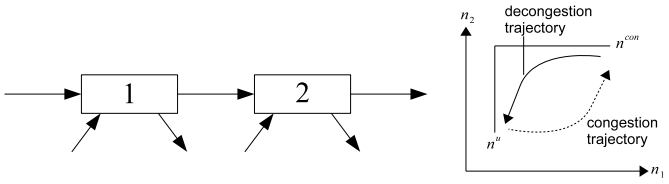


Figure 3.12: By creating the cycle from $n^{con} \rightarrow n^u \rightarrow n^{con}$ a ramp metering strategy can increase off-ramp discharge.

3. A '*free lunch*' result lurks behind Theorem 3.4.2. The result can be understood with the help of Figure 3.12 of a two-cell freeway whose equilibrium set E is shown on the right. By Lemma 3.4.2 the flow in cell 1 is larger than the equilibrium flow in the rectangle $\{n^u < p < n^{con}\}$. The 'decongestion' trajectory constructed in Lemma 3.4.2 moves the system from n^{con} to n^u and causes some additional vehicles to leave the freeway from the off-ramp in cell 1. The remainder of the $[(n_1^{con} - n_1^u) + (n_2^{con} - n_2^u)]$ vehicles causing the initial congestion are 'stored' on the on-ramps in cells 1 and 2. Once the cells become uncongested, the ramp metering strategy can now be changed to release the stored vehicles onto the freeway, thereby creating the congestion and moving the state from n^u to n^{con} as indicated by the 'congestion' trajectory in the figure. Since this trajectory is inside $\{n^u < p < n^{con}\}$ there will again be an additional off-ramp flow. Repeating the two-phase decongestion-congestion cycle provides a free lunch.

3.5 Proofs

Proof of Lemma 3.2.1.

Existence: Let $f = f(r)$ be the equilibrium flow. Define

$$n_i^u = (\bar{\beta}_i\, v_i)^{-1} f_i, \quad 1 \le i \le N. \tag{3.65}$$

Then $n_i(k) \equiv n_i^u$ satisfies (3.6), because (3.6) is equivalent to (3.13). Next, because $0 \le f_i \le F_i$ and $F_i = \bar{\beta}_i\, v_i\, n_i^c$ (see (3.1)), $n_i^u = (\bar{\beta}_i\, v_i)^{-1} f_i \le (\bar{\beta}_i\, v_i)^{-1} F_i = n_i^c$. So n^u is uncongested. It remains to prove that n^u is an equilibrium, i.e., satisfies (3.14), which simplifies to (3.11) because n^u is uncongested. From (3.65), $f_i = \bar{\beta}_i\, v_i\, n_i^u$, and since r is feasible, $f_i \le F_i$. So (3.11) holds.

Uniqueness: Suppose $\{0 \le n_i \le n_i^c; 1 \le i \le N\}$ is an equilibrium, i.e., satisfies (3.14)-(3.15). Since $n_i \le n_i^c$, $\bar{\beta}_i\, v_i\, n_i \le \bar{\beta}_i\, v_i\, n_i^c = F_i$, therefore (3.14) reduces to

$$f_i = \min\{\bar{\beta}_i v_i n_i, w_{i+1}(\bar{n}_{i+1} - n_{i+1})\}.$$

If $f_i \ne \bar{\beta}_i v_i n_i$, it must be that $\bar{\beta}_i v_i n_i > w_{i+1}(\bar{n}_{i+1} - n_{i+1}) \ge w_{i+1}(\bar{n}_{i+1} - n_{i+1}^c) = F_i$. This contradicts $\bar{\beta}_i\, v_i\, n_i \le F_i$, hence f_i must equal $\bar{\beta}_i v_i n_i$, so $n_i = n_i^u$. □

Proof of Proposition 3.2.1

Since $n_{i+1} \le n_{i+1}^c$, from (3.11),

$$f_i = \min\{\bar{\beta}_i v_i n_i, F_i\}.$$

Since $n_i > n_i^c$, one has $\bar{\beta}_i v_i n_i > F_i$ from (3.1); and since r is feasible, $F_i \geq f_i$. Hence $f_i = F_i$ and, by (3.13),

$$f_{i-1} = \bar{\beta}_i^{-1} f_i - r_i = \bar{\beta}_i^{-1} F_i - r_i.$$

Again, as $n_i > n_i^c$, (3.14) implies

$$f_{i-1} = \min\{\bar{\beta}_{i+1} v_{i-1} n_{i-1}, F_{i-1} - w_i(n_i - n_i^c), F_{i-1}\} < F_{i-1}.$$

Lastly, if cell $i-k$ is congested, $n_{i-k} > n_{i-k}^c$, hence

$$f_{i-k-1} = \min\{\bar{\beta}_{i-k-1} v_{i-k-1} n_{i-k-1}, F_{i-k-1} + w_{i-k}[n_{i-k}^c - n_{i-k}], F_{i-k-1}\} < F_{i-k-1},$$

and the remainder of the assertion follows. □

Proof of Corollary 3.2.1

If the equilibrium $n = (n_1, \cdots, n_N)$ is uncongested, then $n = n^u$ by Lemma 3.2.1. So suppose there is at least one congested cell. There are two cases to consider. In the first case, cell N is congested. Since $f_N = \min\{\bar{\beta}_N v_N n_N, F_N\} < F_N$ by strict feasibility, so $n_N < n_N^c$, which means cell N is not congested. In the remaining case, there must exist a pair of adjacent cells $i, i+1$ with i congested and $i+1$ uncongested. But then by Proposition 3.2.1, $f_i = F_i$, which contradicts strict feasibility of r. □

Proof of Proposition 3.2.2

Because cell $i-j$ is congested, i.e., $n_{i-j} > n_{i-j}^c$, and $f_{i-j} < F_{i-j}$,

$$\begin{aligned} f_{i-j} &= \min\{\bar{\beta}_{i-j} v_{i-j} n_{i-j}, F_{i-j} - w_{i-j+1}(n_{i-j+1} - n_{i-j+1}^c), F_{i-j}\} \\ &= F_{i-j} - w_{i-j+1}(n_{i-j+1} - n_{i-j+1}^c) < F_{i-j}, \end{aligned}$$

so that $n_{i-j+1} > n^c_{i-j+1}$, i.e., cell $i-j+1$ is congested. The result follows by induction. □

Proof of Proposition 3.2.3

The assertion is true of segment S^{K+1} because downstream of cell N is free flow by assumption. Consider segment S^k. Since

$$f_{I_k} = \min\{\bar{\beta}_{I_k} v_{I_k} n_{I_k}, F_{I_k} - w_{I_k+1}(n_{I_k+1} - n^c_{I_k+1}), F_{I_k}\} = F_{I_k},$$

we must have $n_{I_k+1} \leq n^c_{I_k+1}$, i.e., cell I_k+1 is uncongested, and (3.18) holds. □

Proof of Lemma 3.2.2

Because of (3.24) the equilibrium flows $f_{I_K+1}, \cdots, f_N$ are strictly below capacity and so, by Corollary 3.2.1, $E^{K+1}(r) = \{n^{u,K+1}\}$, and (3.25) follows from (3.65). □

Proof of Lemma 3.2.3

Let $n^k \in E^k(r)$ be an equilibrium. Then (i) follows from Lemma 3.2.1. Next, according to Proposition 3.2.2 there exists j such that cells $I_{k-1}+1, \cdots, j-1$ are uncongested and $j, \cdots, I_k$ are congested. Hence for $i < j-1$,

$$f_i = \min\{\bar{\beta}_i v_i n^k_i, F_i - w_{i+1}(n^k_{i+1} - n^c_{i+1}), F_i\} = \bar{\beta}_i v_i n^k_i,$$

because $n^k_i \leq n^c_i, n^k_{i+1} \leq n^c_{i+1}$, which proves (3.26).

For $i \geq j+1$,

$$f_i = \min\{\bar{\beta}_i v_i n^k_i, F_i - w_{i+1}(n^k_{i+1} - n^c_{i+1}), F_i\} = F_i - w_{i+1}(n^k_{i+1} - n^c_{i+1}),$$

because $n^k_i > n^c_i, n^k_{i+1} > n^c_{i+1}$, which proves (3.27).

Lastly, because $f_{j-1} < F_{j-1}$,

$$f_{j-1} = \min\{\bar{\beta}_{j-1}v_{j-1}n^k_{j-1}, F_{j-1}-w_j(n^k_j-n^c_j), F_{j-1}\} = \min\{\bar{\beta}_{j-1}v_{j-1}n^k_{j-1}, F_{j-1}-w_j(n^k_j-n^c_j)\}$$

Hence either $f_{j-1} = F_{j-1} - w_j(n^k_j - n^c_j)$ and then (3.28) holds, or $f_{j-1} = \bar{\beta}_{j-1}v_{j-1}n^k_{j-1}$ and then (3.29) holds. □

Proof of Lemma 3.3.1

Suppose $x \leq y$. We must show

$$g_i(x_{i-1}, x_i, x_{i+1}) \leq g_i(y_{i-1}, y_i, y_{i+1}). \tag{3.66}$$

We verify the inequality one coordinate at a time. Suppose first that $x_{i+1} \leq y_{i+1}$ but $x_{i-1} = y_{i-1}, x_i = y_i$. Then from (3.6)-(3.7)

$$\begin{aligned} &g_i(x_{i-1}, x_i, x_{i+1}) - g_i(y_{i-1}, y_i, y_{i+1}) \\ &= -\bar{\beta}_i^{-1}\min\{\bar{\beta}_i v_i x_i, w_{i+1}[\bar{n}_{i+1} - x_{i+1}], F_i\} + \bar{\beta}_i^{-1}\min\{\bar{\beta}_i v_i y_i, w_{i+1}[\bar{n}_{i+1} - x_{i+1}], F_i\} \leq 0. \end{aligned}$$

Suppose next that $x_{i-1} \leq y_{i-1}$ but $x_i = y_i, x_{i+1} = y_{i+1}$. Then from (3.6)-(3.7)

$$\begin{aligned} &g_i(x_{i-1}, x_i, x_{i+1}) - g_i(y_{i-1}, y_i, y_{i+1}) \\ &= \bar{\beta}_{i-1}^{-1}\min\{\bar{\beta}_{i-1}v_{i-1}x_{i-1}, w_i[\bar{n}_i - x_i], F_{i+1}\} - \bar{\beta}_{i-1}^{-1}\min\{\bar{\beta}_{i-1}v_{i-1}y_{i-1}, w_i[\bar{n}_i - x_i], F_{i-1}\} \leq 0. \end{aligned}$$

Lastly suppose $x_i \leq y_i$ but $x_{i-1} = y_{i-1}, x_{i+1} = y_{i+1}$. To show (3.66) consider three separate cases.

Case 1: $x_i < y_i \leq n_i^c$. Then from (3.6)-(3.7) and (3.11)

$$g_i(x_{i-1}, x_i, x_{i+1}) - g_i(y_{i-1}, y_i, y_{i+1})$$

$$= x_i - \bar{\beta}_i^{-1} \min\{\bar{\beta}_i v_i x_i, w_{i+1}[\bar{n}_{i+1} - x_{i+1}], F_i\} - y_i + \bar{\beta}_i^{-1} \min\{\bar{\beta}_i v_i y_i, w_{i+1}[\bar{n}_{i+1} - x_{i+1}], F_i\}$$

$$\begin{cases} = x_i - y_i & \text{if } \bar{\beta}_i v_i x_i \geq \min\{\bar{\beta}_i v_i x_i, w_{i+1}[\bar{n}_{i+1} - x_{i+1}], F_i\} \\ \leq (1 - v_i)x_i - (1 - v_i)y_i & \text{if } \bar{\beta}_i v_i x_i = \min\{\bar{\beta}_i v_i x_i, w_{i+1}[\bar{n}_{i+1} - x_{i+1}], F_i\} \end{cases}$$

≤ 0, because $0 < v_i < 1$.

Case 2: $x_i \leq n_i^c < y_i$. Then from (3.6)-(3.7) and (3.11)

$$g_i(x_{i-1}, x_i, x_{i+1}) - g_i(y_{i-1}, y_i, y_{i+1})$$

$$= x_i - \bar{\beta}_i^{-1} \min\{\bar{\beta}_i v_i x_i, w_{i+1}[\bar{n}_{i+1} - x_{i+1}], F_i\} + \min\{\bar{\beta}_{i-1} v_{i-1} x_{i-1}, w_i[\bar{n}_i - x_i], F_i\}$$

$$- y_i + \bar{\beta}_i^{-1} \min\{\bar{\beta}_i v_i y_i, w_{i+1}[\bar{n}_{i+1} - x_{i+1}], F_i\} - \min\{\bar{\beta}_{i-1} v_{i-1} x_{i-1}, w_i[\bar{n}_i - y_i], F_i\}$$

If $\bar{\beta}_i v_i x_i \geq \min\{\bar{\beta}_i v_i x_i, w_{i+1}[\bar{n}_{i+1} - x_{i+1}], F_i\}$,

$$g_i(x_{i-1}, x_i, x_{i+1}) - g_i(y_{i-1}, y_i, y_{i+1})$$

$$= x_i - y_i + \min\{\bar{\beta}_{i-1} v_{i-1} x_{i-1}, w_i[\bar{n}_i - x_i], F_i\} - \min\{\bar{\beta}_{i-1} v_{i-1} x_{i-1}, w_i[\bar{n}_i - y_i], F_i\}$$

$$\begin{cases} \leq x_i - y_i \leq 0, \text{ if } w_i[\bar{n}_i - x_i] \leq \min\{\bar{\beta}_{i-1} v_{i-1} x_{i-1}, F_i\} \\ = (1 - w_i)(x_i - y_i) \leq 0, \text{ if } w_i[\bar{n}_i - x_i] = \min\{\bar{\beta}_{i-1} v_{i-1} x_{i-1}, F_i\}, \text{ because } 0 < v_i < 1, \end{cases}$$

and if $\bar{\beta}_i v_i x_i < \min\{\bar{\beta}_i v_i x_i, w_{i+1}[\bar{n}_{i+1} - x_{i+1}], F_i\}$,

$$g_i(x_{i-1}, x_i, x_{i+1}) - g_i(y_{i-1}, y_i, y_{i+1})$$

$$\leq x_i - y_i + \min\{\bar{\beta}_{i-1} v_{i-1} x_{i-1}, w_i[\bar{n}_i - x_i], F_i\} - \min\{\bar{\beta}_{i-1} v_{i-1} x_{i-1}, w_i[\bar{n}_i - y_i], F_i\}$$

$$\leq 0, \text{ as before.}$$

Case 3: $n_i^c \le x_i < y_i \le n_i^c$. Then from (3.6)-(3.7) and (3.11)

$$
\begin{aligned}
& g_i(x_{i-1}, x_i, x_{i+1}) - g_i(y_{i-1}, y_i, y_{i+1}) \\
& = x_i - y_i + \min\{\bar{\beta}_{i-1} v_{i-1} x_{i-1}, w_i[\bar{n}_i - x_i]\} - \min\{\bar{\beta}_{i-1} v_{i-1} x_{i-1}, w_i[\bar{n}_i - y_i]\} \\
& \begin{cases} = x_i - y_i \le 0, \text{ if } w_i[\bar{n}_i - y_i] > \bar{\beta}_{i-1} v_{i-1} x_{i-1} \\ \le (1 - w_i)(x_i - y_i) \le 0, \text{ if } w_i[\bar{n}_i - y_i] \le \bar{\beta}_{i-1} v_{i-1} x_{i-1}, \text{ because } 0 < v_i < 1. \end{cases}
\end{aligned}
$$

Thus g is strictly monotone, because if $x \le y$,

$$g_i(x_{i-1}, x_i, x_{i+1}) \le g_i(y_{i-1}, x_i, x_{i+1}) \le g_i(y_{i-1}, y_i, x_{i+1}) \le g_i(y_{i-1}, y_i, y_{i+1});$$

moreover, it is trivial to check that if $x \ne y$ then $g(x) \ne g(y)$.

Lastly g is not strongly monotone, because if $x < y$ but $x_{i-1} = y_{i-1}, x_i = y_i, x_{i+1} = y_{i+1}$, then $g_i(x) = g_i(y)$. □

Proof of Lemma 3.3.2

(i) Since $\hat{n}(0) \le n(0) \le \bar{n}(0)$, monotonicity implies $\hat{n}(1) \le n(1) \le \bar{n}(1)$, and then (3.45) follows by induction.

(ii) Since $\hat{n}(1) \ge \hat{n}(0) = 0$, monotonicity implies $\hat{n}(2) = g(\hat{n}(1)) \ge g(\hat{n}(0)) = \hat{n}(1)$. By induction, the trajectory is increasing: $\hat{n}(k+1) \ge \hat{n}(k)$. Since the trajectory is bounded above by the jam density, it must converge to some equilibrium point, say $\hat{n}^e$. Furthermore, since $n(k) \equiv n^u$ is also a trajectory, by (3.45) one must have $\hat{n}^e \le n^u$, and so (3.44) implies $\hat{n}^e = n^u$.

(iii) Since $\bar{n} = \bar{n}(0) \ge \bar{n}(1)$, monotonicity implies that the trajectory is decreasing: $\bar{n}(k+1) \le \bar{n}(k)$. Since the trajectory is bounded below by 0, it must converge to an equilibrium, say

$\bar{n}^e$. As $n(k) \equiv n^{con}$ is also a trajectory, by (3.45) one must have $\bar{n}^e \geq n^{con}$, and so (3.44) implies $\bar{n}^e = n^{con}$. □

Proof of Lemma 3.3.3

Evaluate the three alternatives in $f_i(p) = \min\{\bar{\beta}_i v_i p_i, w_{i+1}[\bar{n}_{i+1} - p_{i+1}], F_i\}$:

$$\begin{aligned} f_i(p) &= \bar{\beta}_i v_i p_i \geq \bar{\beta}_i v_i n_i^u = \phi_i, \text{ by (3.65), or} \\ &= w_{i+1}[\bar{n}_{i+1} - p_{i+1}] \geq w_{i+1}[\bar{n}_{i+1} - n_{i+1}^{con}] = \phi_i, \text{ by (3.30), or} \\ &= F_i \geq \phi_i, \text{ always.} \end{aligned}$$

Hence $f_i(p) \geq \phi_i$ and assertion (i) follows since for bottleneck cells $\phi_i = F_i$. (ii) follows from the observations that $\hat{n}(k) \leq n(k) \leq \bar{n}(k)$ and $\hat{n}(k) \to n^u, \bar{n}(k) \to n^{con}$ by Lemma 3.3.2; hence every limit point p of $\{n(k)\}$ satisfies $n^u \leq p \leq n^{con}$. To prove (iii) consider a subsequence $\{k_m\}$ along which $f_i(k_m) \to \liminf f_i(k)$ and $n(k_m) \to p \in \omega(n)$. Then $\liminf f_i(k) = f_i(p) \geq \phi_i$, by (i) and (ii). □

Proof of Lemma 3.3.4

Let $p^0 \in \omega(n)$ and $p(k), k \geq 0$, the trajectory starting at p^0. Rewrite (3.6) in terms of this trajectory, and (3.48) as

$$\begin{aligned} p_i(k+1) &= p_i(k) - \bar{\beta}_i^{-1} f_i(p(k)) + f_{i-1}(p(k)) + r_i \\ r_i &= \bar{\beta}_i^{-1}\phi_i - \phi_{i-1}. \end{aligned}$$

Adding these together gives

$$p_i(k+1) = p_i(k) + \bar{\beta}_i^{-1}[\phi_i - f_i(p(k))] - [\phi_{i-1} - f_{i-1}(p(k))].$$

Summing this equation for $i = 2, \cdots, j$ and using (3.52) leads to

$$\begin{aligned}\sum_{i=2}^{j} p_i(k+1) &= \sum_{i=2}^{j} p_i(k) + \sum_{i=2}^{j} \bar{\beta}_i^{-1}[\phi_i - f_i(p(k))] - \sum_{i=2}^{j}[\phi_{i-1} - f_{i-1}(p(k))] \\ &= \sum_{i=2}^{j} p_i(k) + [\phi_j - f_j(p(k))] + \sum_{i=2}^{j} \beta_i \bar{\beta}_i^{-1}[\phi_i - f_i(p(k))].\end{aligned}$$

By Lemma 3.3.3, and taking $j = N$, shows that $\sum_2^N p_i(k)$ is decreasing, and since it is positive, it converges. Hence $f_i(p(k)) \to \phi_i$ for each i. So if $p^1 \in \omega(p^0)$,

$$f(p^1) = \phi, \text{ i.e., } p^1 \in E(r),$$

from which the assertion follows since $p^1 \in \omega(p^0) \subset \omega(n)$, because $\omega(n)$ is invariant. □

Proof of Lemma 3.3.5

By Lemma 3.3.2, $\hat{n}(k) \leq n(k) \leq \bar{n}(k)$, $\hat{n}(k) \to n^u$, $\bar{n}(k) \to n^{con}$. Then (3.54) follows because, by (3.49), $n^{con,K+1} = n^{u,K+1}$. □

Proof of Lemma 3.3.6

Consider (3.53) for $m \geq 1$. By Theorem 3.2.1 $n^{e,m} \in E^m_{I_m - j}(r)$ for some $j \geq 0$, so that cells $I_m - j, \cdots, I_m$ are congested and $I_{m-1} + 1, \cdots, I_m - j - 1$ are uncongested as indicated in Figure 3.13.

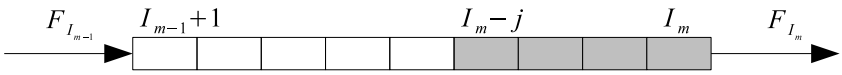


Figure 3.13: In equilibrium $n^{e,m}$ cells $I_m - j, \cdots, I_m$ of S^m are congested.

We will prove (3.53) for m, separately analyzing the three cases: $j = I_{m-1} + 1$, $j = I_m$, and $I_{m-1} + 1 < j < I_m$.

Case (i): $j = I_{m-1} + 1$. In this case at $n^{e,m}$ the entire segment S^m is congested. The induction hypothesis is not used for this case.

By Theorem 3.2.1 $n^{e,m} = n^{con,m}$ and, by (3.30),

$$n_i^{e,m} = n_i^c + w_i^{-1}(F_{i-1} - \phi_{i-1}) > n_i^c, \quad i \in S^m. \tag{3.67}$$

By (3.51) for $\eta > 0$ we can select $\delta > 0$ so that

$$|n^m - n^{e,m}| < \delta \Rightarrow 0 \leq F_{I_m} - f_{I_m}(k) = \eta(k) < \eta. \tag{3.68}$$

Assume for now that

$$n_i^m(k) > n_i^c, \quad k \geq 0,\ i \in S^m, \tag{3.69}$$

so that

$$\begin{aligned} f_i(k) &= \min\{\bar{\beta}_i v_i n_i^m(k), F_i - w_{i+1}[n_{i+1}^m(k) - n_{i+1}^c], F_i\} \\ &= F_i - w_{i+1}[n_{i+1}^m(k) - n_{i+1}^c],\ i = I_{m-1}, \cdots, I_m. \end{aligned} \tag{3.70}$$

Substituting (3.67), (3.68), (3.70), and (3.48) in

$$\begin{aligned} n_i^m(k+1) &= n_i^m(k) - \bar{\beta}_i^{-1} f_i(k) + f_{i-1}(k) + r_i \\ &= n_i^m(k) - \bar{\beta}_i^{-1}[f_i(k) - \phi_i] + f_{i-1}(k) - \phi_{i-1}, \end{aligned}$$

gives, for $i = I_m$,

$$\begin{aligned} n_i^m(k+1) &= n_i^m(k) - \bar{\beta}_i^{-1}[F_{I_m} - \eta(k) - \phi_i] + F_{i-1} - w_i[n_i^m(k) - n_i^c] - \phi_{i-1} \\ &= n_i^m(k) - w_i[n_i^m(k) - n_i^c - w_i^{-1}(F_{i-1} - \phi_{i-1})] + \bar{\beta}_i^{-1}\eta(k), \text{ as } F_{I_m} = \phi_{I_m} \\ &= n_i^m(k) - w_i[n_i^m(k) - n_i^{e,m}] + \bar{\beta}_i^{-1}\eta(k); \end{aligned} \tag{3.71}$$

and, for $i = I_{m-1}+1, \cdots, I_m - 1$,

$$\begin{aligned} n_i^m(k+1) &= n_i^m(k) - \bar{\beta}_i^{-1}[F_i - w_{i+1}(n_{i+1}^m(k) - n_{i+1}^c) - \phi_i] + F_{i-1} - w_i[n_i^m(k) - n_i^c] - \phi_{i-1} \\ &= n_i^m(k) + \bar{\beta}_i^{-1} w_i[n_{i+1}^m(k) - n_{i+1}^c - w_i^{-1}(F_i - \phi_i)] \\ &\quad -w_i[n_i^m(k) - n_i^c - w_i^{-1}(F_{i-1} - \phi_{i-1})] \\ &= n_i^m(k) + \bar{\beta}_i^{-1} w_{i+1}[n_i^m(k) - n_i^{e,m}] - w_i[n_i^m(k) - n_i^{e,m}]. \end{aligned} \tag{3.72}$$

Define the vectors $x^m(k)$ with components $x_i^m(k) = n_i^m(k) - n_i^{e,m}, i \in S^m$. In terms of $x^m(k)$ the difference equations (3.71)-(3.72) can be written as

$$x^m(k+1) = \begin{bmatrix} 1 - w_{I_{m-1}+1} & \beta_{I_{m-1}+1}^{-1} w_{I_{m-1}+2} & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdots & 0 & 1 - w_{I_m - 1} & \beta_{I_m-1}^{-1} w_{I_m} \\ 0 & \cdots & 0 & 0 & 1 - w_{I_m} \end{bmatrix} x^m(k) + \begin{bmatrix} 0 \\ \cdot \\ 0 \\ \beta_{I_m}^{-1}\eta(k) \end{bmatrix}. \tag{3.73}$$

The difference equation (3.73) is of the form

$$x^m(k+1) = Ax^m(k) + u(k), \quad x^m(0) = n^m - n^{e,m},$$

and has the solution

$$x^m(k) = A^k(n^m - n^{e,m}) + \sum_{l=0}^{k-1} A^{k-1-l} u(l).$$

The eigenvalues of A are $(1 - w_{I_{m-1}+1}), \cdots, (1 - w_{I_m})$, all of which lie in $(0, 1)$, since $0 < w_i < 1$. Hence $\| A^k \| \leq M\lambda^k$ for some $M < \infty$ and $0 < \lambda < 1$. Also $|u(l)| \leq (\bar{\beta}_{I_m})^{-1}\eta$. It follows that if $|n^m - n^{e,m}| < \delta$, sufficiently small, then (3.69) holds and $|x^m(k)| \leq \epsilon$ for all $k \geq 0$.

Case (ii): $j = I_m$. In this case cells $j-1 = I_{m-1}+1, \cdots, I_m - 1$ are not congested; so $\phi_j = F_j$, $\phi_i < F_i$, $i < j, i \in S^m$.

By the induction hypothesis, for $\epsilon > 0$ there is $\delta > 0$ such that for $|n - n^e| < \delta$, there is an equilibrium $\tilde{n}^e$ such that

$$|n^{m-1}(k) - \tilde{n}^{e,m-1}| < \epsilon, \ k \geq 0.$$

By Proposition 3.2.3, $\tilde{n}^{e,m-1}_{j+1} < n^c_{j+1}$; hence, for $\epsilon > 0$ small

$$n^{m-1}_{j+1}(k) < \tilde{n}^{e,m-1}_{j+1} + \epsilon < n^c_{j+1}. \tag{3.74}$$

Next, by (3.51) we can select $\delta > 0$ so that

$$|n - n^e| < \delta \Rightarrow 0 \leq F_{I_m - 1} - f_{I_m - 1}(k) = \eta(k) \to 0. \tag{3.75}$$

By Lemma 3.2.3, $n^{e,m}$ has the form (see bottom part of Figure 3.3)

$$n^{e,m}_i = \begin{cases} n^c_j + (1-2\psi) w_j^{-1}(F_{j-1} - \phi_{j-1}), & i = j, \text{ for some } 0 \leq \psi \leq 1/2 \\ n^u_i = (\bar{\beta}_i v_i)^{-1} \phi_i < n^c_i, & i < j, i \in S^m \end{cases}.$$

We now examine the trajectory $\{n(k)\}$ starting at n. Assume for now that

$$n^m_j(k) \ < \ n^{e,m}_j + \psi w_j^{-1}(F_{j-1} - \phi_{j-1}) = n^c_j + (1-\psi) w_j^{-1}(F_{j-1} - \phi_{j-1}), \tag{3.76}$$

$$n^m_i(k) \ < \ n^{e,m}_i + \psi (\bar{\beta}_i v_i)^{-1}(n^c_i - n^{e,m}_i), \ i < j, i \in S^m. \tag{3.77}$$

From (3.74) and (3.76)

$$\begin{aligned} f_j(k) &= \min\{\bar{\beta}_j v_j n^m_j(k), F_j - w_{j+1}(n^{m-1}_{j+1}(k) - n^c_{j+1}), F_j\} \\ &= \min\{\bar{\beta}_j v_j n^m_j(k), F_j\} = \begin{cases} F_j, & \text{if } n^m_j(k) > n^c_j \\ \bar{\beta}_j v_j n^m_j(k), & \text{if } n^m_j(k) \leq n^c_j \end{cases}. \end{aligned} \tag{3.78}$$

Next,

$$f_{j-1}(k) = \min\{\bar{\beta}_{j-1}v_{j-1}n^m_{j-1}(k), F_{j-1} - w_j(n^m_j(k) - n^c_j), F_{j+1}\}. \tag{3.79}$$

From (3.77),

$$\bar{\beta}_{j-1}v_{j-1}n^m_{j-1}(k) < \bar{\beta}_{j-1}v_{j-1}[n^{e,m}_{j-1} + \psi(\bar{\beta}_{j-1}v_{j-1})^{-1}(n^c_{j-1} - n^{e,m}_{j-1})] = \phi_{j-1} + \psi(F_{j-1} - \phi_{j-1}),$$

and from (3.76)

$$F_{j-1} - w_j(n^m_j(k) - n^c_j) \geq F_{j-1} - w_j[(1-\psi)w_j^{-1}(F_{j-1} - \phi_{j-1})] = \psi F_{j-1} + (1-\psi)\phi_{j-1}.$$

Substituting the preceding two inequalities into (3.79) gives

$$f_{j-1}(k) = \bar{\beta}_{j-1}v_{j-1}n^m_{j-1}(k). \tag{3.80}$$

Lastly, for $i \leq j-2,\ i \in S^m$, from (3.77)

$$f_i(m) = \min\{\bar{\beta}_i v_i n^m_i(k), F_i - w_{i+1}(n^m_{i+1}(k) - n^c_{i+1}, F_i\} = \bar{\beta}_i v_i n^m_i(k). \tag{3.81}$$

Substituting (3.78), (3.80), (3.81), and (3.48) in

$$\begin{aligned} n^m_i(k+1) &= n^m_i(k) - \bar{\beta}_i^{-1} f_i(k) + f_{i-1}(k) + r_i \\ &= n^m_i(k) - \bar{\beta}_i^{-1}[f_i(k) - \phi_i] + [f_{i-1}(k) - \phi_{i-1}], \end{aligned}$$

gives the difference equation system for $\{n^m(k)\}$:

$$\begin{aligned} n^m_j(k+1) &= n^m_j(k) - \bar{\beta}_j^{-1}[\min\{\bar{\beta}_j v_j n^m_j(k), F_j\} - \bar{\beta}_j v_j n^{e,m}_j] \\ &\quad + \bar{\beta}_{j-1}v_{j-1}[n^m_{j-1}(k) - n^{e,m}_{j-1}(k)] \\ n^m_i(k+1) &= n^m_i(k) - v_i[n^m_i(k) - n^{e,m}_i(k)] + \bar{\beta}_{i-1}v_{i-1}[n^m_{i-1} - n^{e,m}_{i-1}], \\ &\quad i = I_{m-1} + 2, \cdots, j-1 \\ n^m_{I_{m-1}+1}(k+1) &= n^m_{I_{m-1}+1}(k) - v_{I_{m-1}+1}(k)[n^m_{I_{m-1}+1}(k) - n^{e,m}_{I_{m-1}+1}(k)] - \eta(k). \end{aligned}$$

In terms of the variables $x_i^m(k) = n_i^m(k) - n_i^{e,m},\ i \leq j, i \in S^m$, this system can be rewritten as

$$x_j^m(k+1) = x_j^m(k) - \bar{\beta}_j^{-1}[\min\{\bar{\beta}_j v_j n_j^m(k), F_j\} - \bar{\beta}_j v_j n_j^{e,m}] + \bar{\beta}_{j+1} v_{j+1} x_{j+1}^m(k) \tag{3.82}$$

and

$$\begin{bmatrix} x_{I_{m-1}+1}(k+1) \\ \cdot \\ x_{j-2}(k+1) \\ x_{j-1}(k+1) \end{bmatrix} = \begin{bmatrix} 1 - v_{I_{m-1}+1} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdots & \bar{\beta}_{j-3} v_{j-3} & 1 - v_{j-2} & 0 \\ 0 & \cdots & 0 & \bar{\beta}_{j-2} v_{j-2} & 1 - v_{j-1} \end{bmatrix} \begin{bmatrix} x_{I_{m-1}+1}(k) \\ \cdot \\ x_{j-2}(k) \\ x_{j-1}(k) \end{bmatrix} + \begin{bmatrix} -\eta(k) \\ 0 \\ \cdot \\ 0 \end{bmatrix}. \tag{3.83}$$

The difference equation (3.83) is of the form

$$z(k+1) = Az(k) - b\eta(k),$$

and has the solution

$$z(k) = A^k z(0) - \sum_{l=0}^{k-1} A^{k-1-l} b\eta(l).$$

The eigenvalues of A are $(1 - v_{I_{m-1}+1}), \cdots, (1 - v_{j-1})$, all of which lie in $(0, 1)$, since $0 < v_i < 1$. By (3.75), $\eta(l) \to 0,\ \eta(l) \geq 0$. Hence $z(k) \to z^* \leq 0$, and (3.77) is assured by induction. Furthermore, because $\eta(l) \geq 0$,

$$x_{j-1}^m(k) \leq M\lambda^k, \tag{3.84}$$

for some $M < \infty$ and $0 < \lambda < 1$.

Lastly, rewrite (3.82) as

$$x_j^m(k+1) = \begin{cases} (1-v_j)x_j^m + \bar{\beta}_{j-1}v_{j-1}x_{j-1}^m(k), & \text{if } x_j^m \leq -\Delta \\ x_j^m(k) + \Delta + \bar{\beta}_{j-1}v_{j-1}x_{j-1}^m(k), & \text{if } x_j^m \geq -\Delta \end{cases}, \tag{3.85}$$

in which $\Delta = (1-2\psi)w_j^{-1}(F_{j-1} - \phi_{j-1})$.

Because $\Delta > 0$ and $x_{j-1}^m \to 0$, the second alternative in (3.85) cannot hold for $k \geq K$, for some finite K, and so

$$x_j^m(k) = (1-v_j)^{k-K}x_j^m(K) + \sum_{l=K}^{k-1}(1-v_j)^{k-1-l}\bar{\beta}_{j-1}v_{j-1}x_{j-1}^m(l),$$

can be made arbitrarily small, proving (3.76).

Case (iii): $I_{m-1} + 1 < j < I_m$. In this case at $n^{e,m}$ cells $j, \cdots, I_m$ are congested and cells $I_{m-1}+1, \cdots, j-1$ are uncongested. The proof for this case combines the argument in Case (i) for the congested cells and the argument in Case (ii) for the uncongested cells. The details are omitted. □

Proof of Corollary 3.3.1

By Theorem 3.2.1 $E(r)$ is a K-dimensional manifold. By Theorem 3.3.3

$$\Sigma = \bigcup_{n^e \in E(r)} \Sigma(n).$$

By Lemma 3.3.2 every trajectory starting at $n \leq n^u$ converges to n^u and every trajectory starting at $n \geq n^{con}$ converges to n^{con}. Because $E(r)$ is ordered, it is not very difficult

to show, using monotonicity, that the stable manifolds of all equilibria $n^e \neq n^u, n^{con}$ are diffeomorphic. The assertion then follows. □

Proof of Lemma 3.4.1

First,

$$f_N(p) = \min\{\bar{\beta}_N v_N p_N, F_N\} \geq \min\{\bar{\beta}_N v_N n_N^u, F_N\} = \phi_N = F_N.$$

Next, for $j \leq i < N$ evaluate the three terms in $f_i(p) = \min\{\bar{\beta}_i v_i p_i, F_i - w_{i+1}[\bar{n}_{i+1} - p_{i+1}], F_i\}$ gives

$$\begin{aligned} f_i(p) &= \bar{\beta}_i v_i p_i > \bar{\beta}_i v_i n_i^u = \phi_i, \text{ or} \\ &= F_i - w_{i+1}[\bar{n}_{i+1} - p_{i+1}] > F_i - w_{i+1}[\bar{n}_{i+1} - n_{i+1}^{con}] = \phi_i, \text{ or} \\ &= F_i > \phi_i, \end{aligned}$$

so $f_i(p) > \phi_i$. The last clause in (3.61) follows from $n_i^u = p_i,\ i > j$. □

Proof of Lemma 3.4.2

Set $\mu_i(k) \equiv 0,\ i < j$. Following (3.6), the controlled trajectory is given by

$$p_i(k+1) = p_i(k) - \bar{\beta}_i^{-1} f_i(p(k) + f_{i-1}(p(k)) + r_i - \mu_i(k),\ 1 \leq i \leq N, \quad k \geq 0. \tag{3.86}$$

Observe that for $i < j$, $p_i(0) = n_i^e = n_i^u$ and $r_i(k) = r_i$. Hence under any metering strategy of the form (3.62), $p_i(k) \equiv n_i^e,\ i < j$. Thus the metering strategy affects the densities only in cells $j, \cdots, N$.

Rewrite (3.86) as

$$p_i(k+1) = g_i(p(k)) - \mu_i(k),\ 1 \leq i \leq N, \quad k \geq 0,$$

and define the metering strategy by

$$\mu_i(k) = \begin{cases} \rho, & \text{if } g_i(p(k)) \geq n_i^u + \rho \\ g_i(p(k)) - n_i^u, & \text{if } n_i^u \leq g_i(p(k)) < n_i^u + \rho \end{cases} . \tag{3.87}$$

Since $r_i - \mu_i(k) \geq r_i(k) - \rho \geq 0$, the metering strategy is feasible (on-ramp flows are non-negative). By construction of μ, $n^u \prec p(k)$. By monotonicity, if $p(k) \prec n^e$ then $p(k+1) = g(p(k)) - \mu(k) \prec g(n^e) - \mu(k)$, so (3.63) holds by induction.

We now prove (3.64). Recall that

$$r_i = \bar{\beta}_i^{-1}\phi_i - \phi_{i-1},$$

and substitute for r_i in (3.86) to get

$$p_i(k+1) = p_i(k) + \bar{\beta}_i^{-1}[\phi_i - f_i(p(k)] - [\phi_{i-1} - f_{i-1}(p(k))] - \mu_i(k),\ j \leq i \leq N.$$

Adding these equations gives

$$\begin{aligned} \sum_{i=j0}^{N} p_i(k+1) &= \sum_{i=j}^{N} p_i(k) + \sum_{i=j}^{N} \bar{\beta}_i^{-1}[\phi_i - f_i(p(k)] - \sum_{i=j}^{N}[\phi_{i-1} - f_{i-1}(p(k))] - \sum_{i=j}^{N} \mu_i(k) \\ &= \sum_{i=j}^{N} p_i(k) + \sum_{i=j}^{N} (\bar{\beta}_i^{-1} - 1)[\phi_i - f_i(p(k)] - \sum_{i=j}^{N} \mu_i(k) \\ &\quad + \bar{\beta}_N^{-1}[\phi_N - f_N(p(k)] - [\phi_{j-1} - f_{j-1}(p(k))] \\ &= \sum_{i=j}^{N} p_i(k) + \sum_{i=j}^{N} (\bar{\beta}_i^{-1} - 1)[\phi_i - f_i(p(k)] - \sum_{i=j}^{N} \mu_i(k) < -\sum_{i=j}^{N} \mu_i(k), \end{aligned}$$

because $f_N(p(k)) = \phi_N = F_N$, $f_{j-1}(p(k)) = \phi_{j-1}$ and $\phi_i - f_i(p(k)) < 0$ by (3.61). Moreover, from (3.87), $\sum_{i=j}^{N} \mu_i(k) \geq \rho$ for each k for which $g_i(p(k) \geq n_i^u + \rho$ for some i. It follows that $p(K) = n^u$ for some $K \leq [\sum_i (n_i^e - n_i^u)]/\rho$. □

Chapter 4

CTMSIM: Interactive CTM Simulator for MATLAB

4.1 Motivation

Freeway traffic measurement data are collected by PeMS [1]. In addition to flow, speed and occupancy time contours, PeMS provides the following performance measures:

- *VHT* - Vehicle Hours Traveled, for a given unit of time and a given section of freeway, the amount of time spent by all of the vehicles on the freeway.
- *VMT* - Vehicle Miles Traveled, for a given unit of time and a given section of the freeway, the sum of the miles of freeway driven by each vehicle.

- *Delay* (vehicle-hours) - difference between the actual VHT and the VHT that would be incurred if vehicles traveled at free flow speed. The value is positive when the road is congested, otherwise it is zero.

- *Productivity Loss* (lane-mile-hours) - the number of lane-mile-hours on the freeway lost due to reduced flow, while operating under congested instead of free-flow conditions. The value is positive only in congestion, otherwise it is zero.

Our first requirement for a macroscopic freeway traffic simulator is the ability to compute these quantities and export the computed data for the purpose of report generation and comparison with the values provided by PeMS.

PeMS is a starting point for any freeway study in TOPl. It allows to extract freeway geometry, estimate fundamental diagrams, split ratios for the off-ramps, and generate a demand profile for the on-ramps. All these data belong to a configuration file of a macroscopic simulator. Thus, for the user to be able to generate his/her own configuration files using PeMS data, the format of such files should be transparent and well documented. This is our second requirement.

Ideally, assuming PeMS data were consistent and came from healthy detectors, plugging PeMS on-ramp flow values into the simulator should produce flow, speed density time contours as well as VHT and VMT that match those reported by PeMS. This would validate the traffic flow model used by a simulator.

Another important requirement was for the user to be able to plug in his/her own ramp controllers, develop and test his/her own ramp metering strategies and compare them with

the known ones or with each other.

Finally, we wanted the simulation to be interactive, so the user could see the evolution of the freeway state, pause the simulation and change some parameters like fundamental diagram—to simulate an incident, or switching certain controllers on or off, and then continue. Interactivity also helps to calibrate the model—adjust fundamental diagrams, on-ramp demands and off-ramp split ratios where necessary. On the other hand, it should be able to execute in a batch mode in case the user wanted to run simulations for many configurations automatically.

Having looked at available macroscopic tools (see Section 2.3), the TOPl group made a decision to develop its own simulator based on the CTM model that would satisfy the listed requirements and have an intuitive user interface. CTMSIM [8] is implemented in MATLAB making it useful and easy to handle for transportation researchers who can quickly and seemlessly develop plug-ins and extensions for it in a familiar environment.

4.2 CTMSIM

4.2.1 Computational Model

CTMSIM is based on CTM model described in Chapter 3. Only now the model parameters and variables are more general. They are summarized in Table 4.1. Compare with Table 3.1.

Symbol	Name	Unit
N	number of cells	dimensionless
Δt	sampling period (time step)	hours
Δx_i	cell length	miles
F_i	total cell capacity	vehicles per hour (vph)
R_i	on-ramp capacity	vehicles per hour (vph)
S_i	off-ramp capacity	vehicles per hour (vph)
v_i	free flow speed	miles per hour (mph)
w_i	congestion wave speed	miles per hour (mph)
$\bar{\rho}_i$	jam density	vehicles per mile (vpm)
ρ_i^c	critical density	vehicles per mile (vpm)
β_i	split ratio	$\in [0,1]$, dimensionless
$\bar{\beta}_i$	complementary split ratio $= 1 - \beta_i$	$\in (0,1]$, dimensionless
γ_i	on-ramp flow blending factor	$\in [0,1]$, dimensionless
ξ_i	on-ramp flow allocation factor	$\in [0,1]$, dimensionless
k	period number	dimensionless
$s_i(k), r_i(k)$	off-ramp, on-ramp flow in cell i in period k	vehicles per hour (vph)
$d_i(k)$	on-ramp demand in cell i in period k	vehicles per hour (vph)
$q_i(k)$	on-ramp queue size in cell i in period k	vehicles
$f_i(k)$	flow from cell i to $i+1$ in period k	vehicles per hour (vph)
$\rho_i(k)$	density in cell i in period k	vehicles per mile (vpm)
$V_i(k)$	actual speed in cell i in period k	miles per hour (mph)
$TT(k)$	travel time in period k	hours
$VHT(k)$	vehicle hours traveled in period k	vehicle hours
$VMT(k)$	vehicle miles traveled in period k	vehicle miles
$D(k)$	delay in period k	vehicle hours
$PL(k)$	productivity loss in period k	lane mile hours (lmh)

Table 4.1: Model parameters and variables used in CTMSIM.

The parameter γ_i determines the influence of the on-ramp flow on the mainline flow that enters ith cell. It reflects the position of the on-ramp within the cell, with larger values of γ_i corresponding to on-ramps that are closer to the upstream edge.

The parameter ξ_i determines the allotment of available space to vehicles entering from the on-ramp. It reflects the geometrical layout of the cell. For example, if the on-ramp is located at the midpoint, incoming vehicles will only have access to the downstream half of the cell.

The on-ramp demand $d_i(k)$ is the number of vehicles per unit of time intending to enter freeway at the ith cell, as opposed to on-ramp flow $r_i(k)$—the number of vehicles per unit of time actually entering freeway at this cell.

The initial condition for the system is the N-dimensional vector of densities $\rho(0)$ at time step 0. Given the initial condition, on-ramp demands and off-ramp split ratios, CTMSIM computes the system evolution in time using the following steps.

1. Check if the user-set value of Δt is valid. It must satisfy

$$\Delta t < \min_i \frac{\Delta x_i}{v_i}. \tag{4.1}$$

2. Set time step $k = 0$.

3. Initialize on-ramp queue size $q_i(0) = 0$, $i = 1..N$.

4. Initialize on-ramp flow $r_i(0) = R_i$, $i = 1..N$. In cells without on-ramps, R_i is assumed to be 0.

5. Compute on-ramp flows

$$\begin{aligned} r_i(k+1) \quad = \quad & \min \Big\{ d_i(k+1) + \frac{q_i(k)}{\Delta t}, \\ & \xi_i(\bar{\rho}_i - \rho_i(k)) \frac{\Delta x_i}{\Delta t}, \\ & R_i, \\ & \max\{\mathcal{C}(r_i(k)), \mathcal{Q}(r_i(k))\} \Big\}, \quad i = 1..N, \end{aligned} \tag{4.2}$$

where $\mathcal{C}(r_i(k))$ denotes flow value suggested by on-ramp mainline controller, and $\mathcal{Q}(r_i(k))$ denotes the flow value coming from on-ramp queue controller.

6. Update queue sizes

$$q_i(k+1) = \max\left\{q_i(k) + \left(d_i(k+1) - r_i(k+1)\right)\Delta t, 0\right\}, \quad i = 1..N. \tag{4.3}$$

7. Compute cell-to-cell flows

$$\begin{aligned} f_i(k+1) &= \min\Bigg\{\bar{\beta}_i v_i\left(\rho_i(k) + \gamma_i r_i(k+1)\frac{\Delta t}{\Delta x_i}\right), \\ &\quad w_{i+1}\left(\bar{\rho}_{i+1} - \left(\rho_{i+1}(k) + \gamma_{i+1} r_{i+1}(k+1)\frac{\Delta t}{\Delta x_{i+1}}\right)\right), \\ &\quad \frac{\bar{\beta}_i}{\beta_i} S_i, \\ &\quad F_i\Bigg\}, \quad i = 1..(N-1), \end{aligned} \tag{4.4}$$

$$f_N(k+1) = \min\left\{\bar{\beta}_N v_N\left(\rho_N(k) + \gamma_N r_N(k+1)\frac{\Delta t}{\Delta x_N}\right), \frac{\bar{\beta}_N}{\beta_N} S_N, F_N\right\}. \tag{4.5}$$

8. Compute off-ramp flows

$$s_i(k+1) = \begin{cases} \frac{\beta_i}{\beta_i} f_i(k+1), & \text{if } \beta_i < 1 \\ \min\left\{v_i\left(\rho_i(k) + \gamma_i r_i(k+1)\frac{\Delta t}{\Delta x_i}\right), S_i\right\}, & \text{if } \beta_i = 1 \end{cases}, \quad i = 1..N. \tag{4.6}$$

9. Compute densities

$$\rho_i(k+1) = \rho_i(k) + \frac{\Delta t}{\Delta x_i}\left(f_{i-1}(k+1) + r_i(k+1) - f_i(k+1) - s_i(k+1)\right), \quad i = 1..N. \tag{4.7}$$

10. Compute actual speeds

$$V_i(k+1) = \min\left\{v_i, \frac{f_i(k+1) + s_i(k+1)}{\rho_i(k+1)}\right\}, \quad i = 1..N. \tag{4.8}$$

11. Compute travel time[1]

$$TT(k+1) = \sum_{i=1}^{N} \frac{\Delta x_i}{V_i(k+1)}. \tag{4.9}$$

12. Compute VHT. VHT in cell i:

$$VHT_i(k+1) = (\rho_i(k+1)\Delta x_i + q_i(k+1))\,\Delta t. \tag{4.10}$$

Total VHT:

$$VHT(k+1) = \sum_{i=1}^{N} VHT_i(k+1) \tag{4.11}$$

13. Compute VMT. VMT in cell i:

$$VMT_i(k+1) = \rho_i(k+1)V_i(k+1)\Delta t \Delta x_i. \tag{4.12}$$

Total VMT:

$$VMT(k+1) = \sum_{i=1}^{N} VMT_i(k+1). \tag{4.13}$$

14. Compute delay. Delay in cell i:

$$D_i(k+1) = \begin{cases} 0, & \text{if } \rho_i(k+1) \le \rho_i^c \\ VHT_i - VMT_i/v_i, & \text{if } \rho_i(k+1) > \rho_i^c \end{cases}, \quad i = 1..N. \tag{4.14}$$

Total delay:

$$D(k+1) = \sum_{i=1}^{N} D_i(k+1). \tag{4.15}$$

Note: Total delay is the sum of freeway delay and queuing delay, because of (4.10).

[1]Unless specified otherwise, by travel time we understand *instantaneous* travel time as opposed to *actual* travel time. Instantaneous travel time is the travel time that would occur, if the traffic speed in each cell stayed constant assuming values at the current time step. It does not include time spent in queues. Actual travel time is the travel time computed using actual time-varying traffic speed values and includes time spent in a queue.

15. Compute productivity loss. Productivity loss in cell i:

$$PL_i(k+1) = \begin{cases} 0, & \text{if } \rho_i(k+1) \le \rho_i^c \\ \left(1 - \frac{f_i(k+1)}{F_i}\right) \Delta t \Delta x_i, & \text{if } \rho_i(k+1) > \rho_i^c \end{cases}, \quad i = 1..N. \tag{4.16}$$

Total productivity loss:

$$PL(k+1) = \sum_{i=1}^{N} PL_i(k+1). \tag{4.17}$$

16. Set $k = k + 1$.

17. Go to step 5.

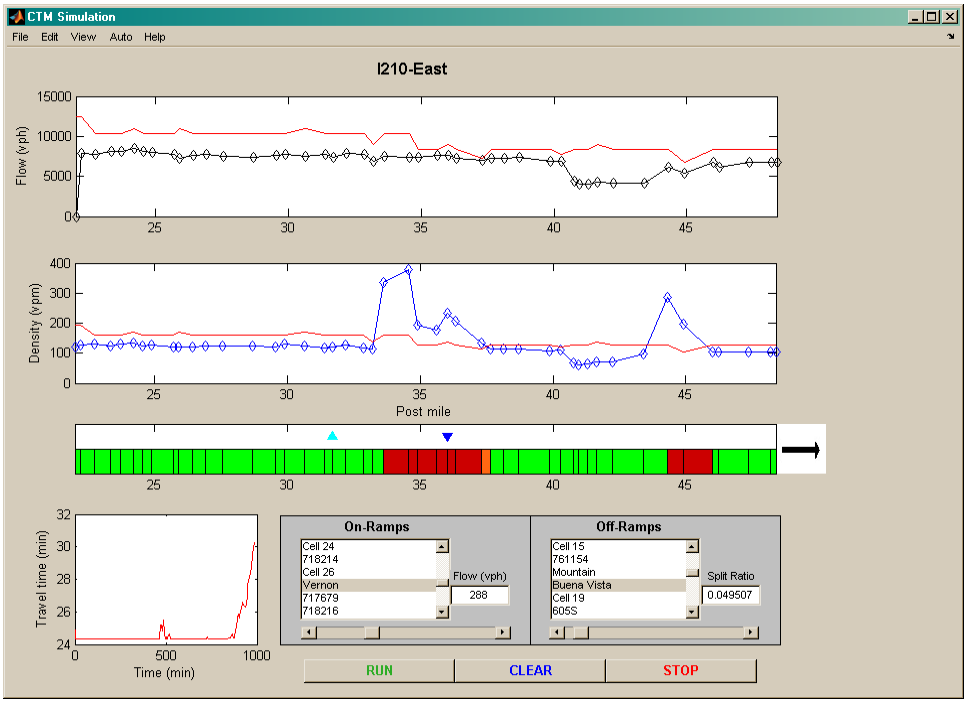
4.2.2 User Interface

The core application of the CTMSIM package is `ctmsim`. It can operate in both graphical and batch mode. To start `ctmsim` in graphical mode, type

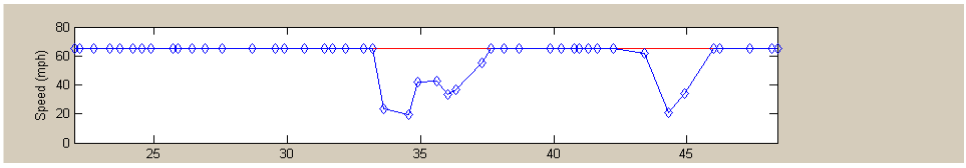
```
>> ctmsim myconfig
```

Here it is assumed that `myconfig.mat` is a configuration file with freeway cell specification, and optional simulation and display parameters, whose format is described in Appendix A. As a result of this command, the main application window pops up. Figure 4.1a presents the look and feel of the `ctmsim` GUI. All the quantities computed in (4.2)-(4.17) can be displayed in the main window as simulation runs. The default display mode is as shown in Figure 4.1a. Plotted in the main area (black and blue) are flows (4.4)-(4.5) and densities (4.7) in each cell at the current time step. The red line represents cell capacities on the flow plot, and critical densities on the density plot. Instead of densities, the actual speeds (4.8)

can be plotted (`View` → `Display Speeds`, see Figure 4.1b). The red line on the speed plot marks free flow speeds.



(a)



(b)

Figure 4.1: (a) Main window of the `ctmsim` application.
(b) Instead of densities, the user may choose to display speeds.

Instead of flows and densities/speeds, the main area of the application window can be used to plot:

- On-ramp demands, flows (4.2) and queue sizes (4.3) (`View` → `On-Ramp Demands and Queues`, Figure 4.2a). Demand values are shown as yellow bars, and flows as blue bars

in each cell. If a user sees only a blue bar in some cell, it means that on-ramp flow equals the demand. If, on the other hand, only the yellow bar is visible, it means that on-ramp flow is zero. Otherwise, the user should see both yellow and blue bars. If the yellow bar is higher than the blue, this indicates that either some control is turned on at this on-ramp, or that the demand exceeds ramp capacity, in which case the queue starts to grow. The situation when the blue bar is higher than the yellow is only possible if there exists a queue at the on-ramp, and the ramp can let through more vehicles than the demand, resulting in decreasing queue size.

- Off-ramp flows (4.6) and split ratios (`View` → `Off-Ramp Flows and Split Ratios`, Figure 4.2b). Split ratios are either constant (user may change them manually as simulation runs), or taken from the split ratio profile (configuration variable `betaProfile`[2]).

- VHT (4.10) and VMT (4.12) (`View` → `VHT and VMT`, Figure 4.2c). At the VHT plot, the delay portion (4.14) is shown in dark red. At the VMT plot, maximum VMT values are marked as empty bars. They are computed by substituting $\rho_i(k+1)$ with ρ_i^c in (4.12) and visualize how well the cell capacities are utilized. The ultimate goal of any ramp metering strategy is to keep the VMT bars at maximum in congestion periods.

Instead of plotting selected quantities only at the current simulation step, the user may choose to see the history of how the system evolved in time by looking at time contours (`View` → `Timeseries Contours`). Figure 4.3 shows an example of flow and speed time contours. It is not recommended to display time contours while running the simulation

[2]All configuration variables are described in Appendix A.

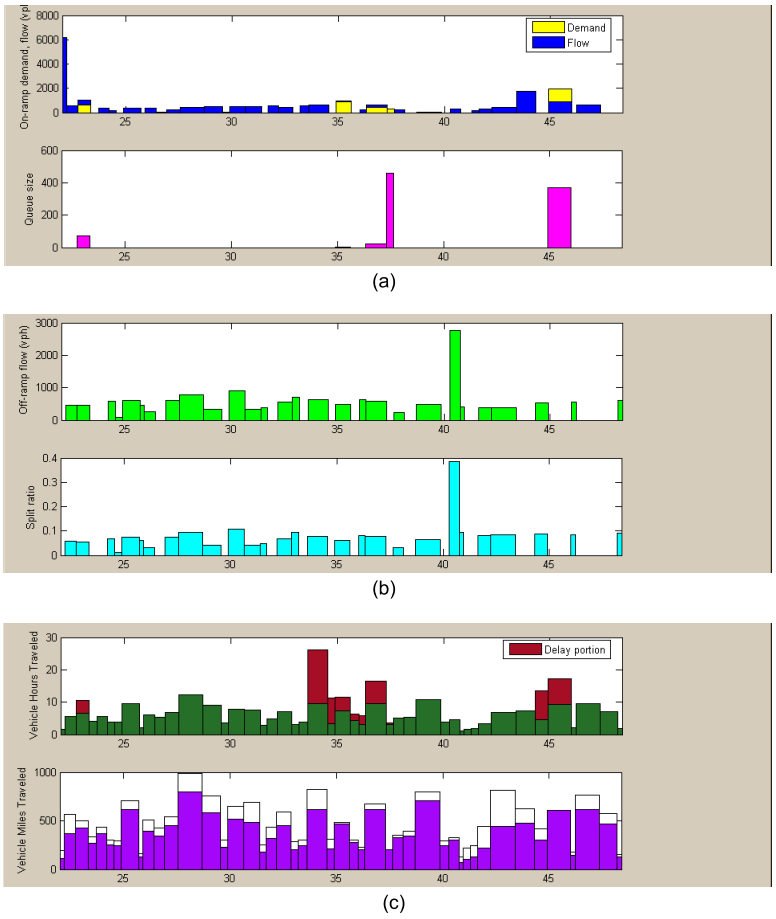


Figure 4.2: Display (a) On-ramp demands, flows and queues.
(b) Off-ramp flows and split ratios.
(c) VHT, delay and VMT per cell.

because drawing the display significantly slows down program execution.

Below the main plotting area (Figure 4.1a) there is a display of the freeway broken up into cells. The arrow specifies the direction of traffic flow: left-to-right when post miles are increasing in the same direction as traffic flows (corresponds to directions north or east on California freeways), or right-to-left when post miles are decreasing in the direction of traffic flow (south or west on California freeways).

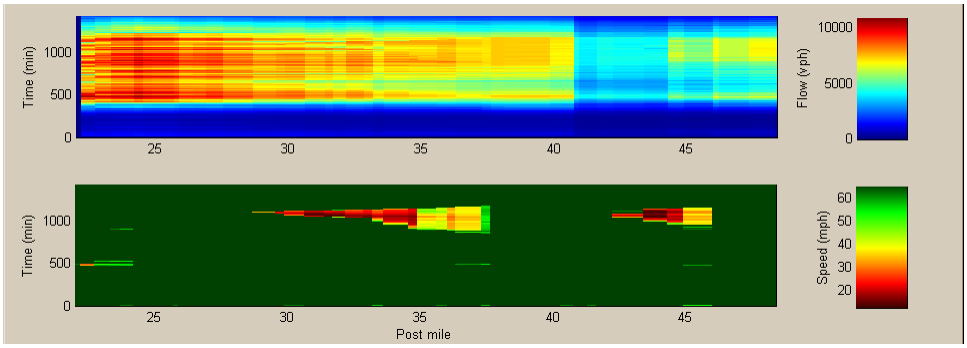


Figure 4.3: Flow and speed time contours.

Color coding is used to display the status of each cell: green means free flow, red means congested. There are two auxiliary colors: yellow indicates free flow at 97% of capacity, orange indicates congested flow at 97% of capacity. The theshold 97% is a configurable parameter (configuration variable `yoColorRatio`).

The blue triangular marker (Figure 4.1) shows which on-ramp is currently selected from the list of on-ramps. The cyan triangular marker shows which off-ramp is currently selected from the list of off-ramps.

Ramp lists are at the bottom of the `ctmsim` window. They may be used to manually change on-ramp flows or off-ramp split ratios, or just to see where a particular ramp is located on the freeway.

The area at the bottom left corner of the `ctmsim` window (Figure 4.1a) is dedicated to plotting aggregate quantities: travel time (4.9), VHT (4.11) (Figure 4.4a), VMT (4.13) (Figure 4.4b), delay (4.15) (Figure 4.4c), and productivity loss (4.17) (Figure 4.4d).

As the simulation runs, all plotted data are updated. The frequency of such updates (once

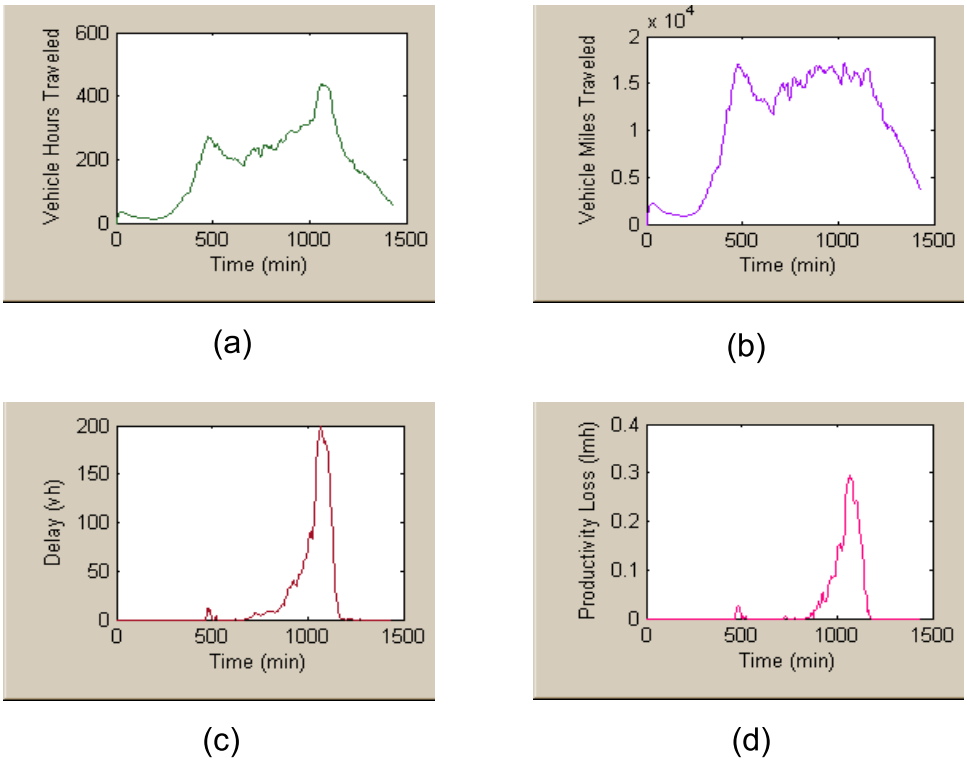


Figure 4.4: Display aggregate data. (a) VHT. (b) VMT. (c) Delay (d) Productivity loss.

per 5 simulation minutes, or once per simulation hour) is defined by the user (configuration variable `plotTS`).

If the demand profile is specified by the user (configuration variable `demandProfile`), the on-ramp demand values can be updated automatically (`Auto` → `On-Ramp Demand`). If auto demand is turned on, the user can toggle the on-ramp control switch by going to `Auto` → `On-Ramp Control`. Once on-ramp control is turned on, the queue control switch can be toggled (`Auto` → `On-Ramp Queue Control`).

Similarly to the demand profile for on-ramps, the user may choose to specify the profile of

split ratios for off-ramps (configuration variable `betaProfile`). To update split ratios by values from the profile, go to `Auto` → `Off-Ramp Split Ratio`. If instead of split ratios, the user wants to specify off-ramp flows, he/she can create a profile of off-ramp flows (configuration variable `frflowProfile`) and run the simulation with automatically updated off-ramp flows instead (`Auto` → `Off-Ramp Flow`). Not more than one of the two profiles, split ratio or off-ramp flow, can be turned on at the same time.

At the bottom of the `ctmsim` window (Figure 4.1) there are `RUN`, `CLEAR` and `STOP` buttons. To start the simulation, press `RUN`. `STOP` button pauses the simulation. At this point the entire simulation run can be saved (`File` → `Save Simulation`). Pressing `RUN` again will resume the simulation from the point it has been stopped. To start from the beginning, before `RUN` press the `CLEAR` button. When pressed while the simulation is running, `CLEAR` button only clears the existing data history without resetting the state to initial conditions and time step to zero. One may choose to stop simulation any time, save it, load it again at a later moment (`File` → `Load Simulation`) and run it from the point it was left off.

A user cannot exit `ctmsim` while the simulation is running. It has to be stopped first. If the simulation was not saved, the user is asked if he/she wishes to save it upon exiting or loading other simulation.

To modify simulation and model parameters (Table 4.1 entries that do not depend on time step k), several editor windows can be called directly from `ctmsim`.

- Settings editor (`Edit` → `Settings`, Figure 4.5). It is used to modify general simulation parameters.

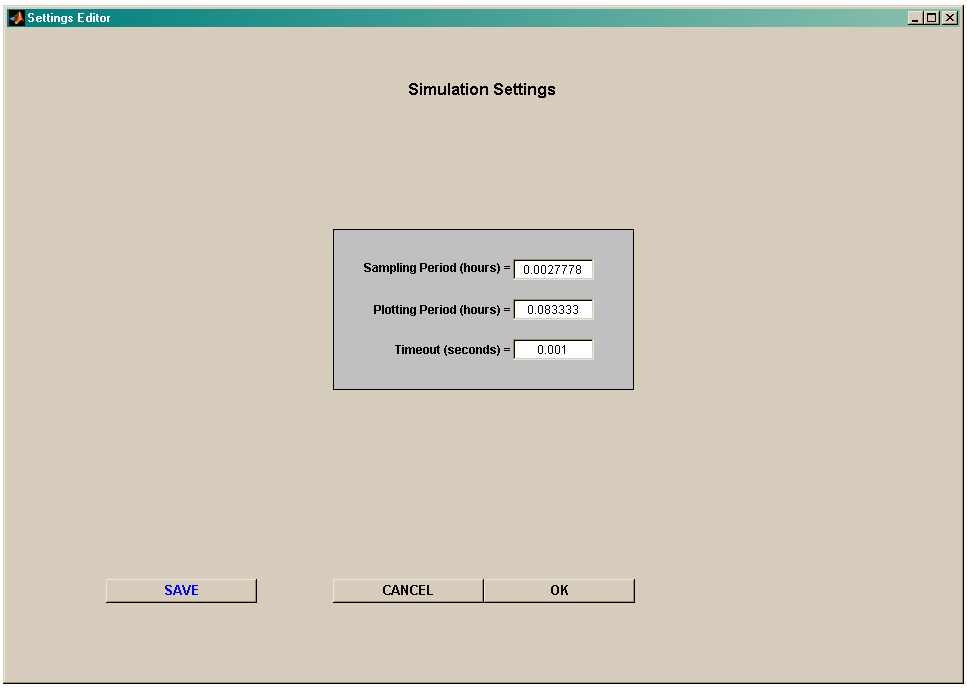


Figure 4.5: Editor for simulation parameters.

Sampling period Δt (configuration variable `TS`)—user value must satisfy 4.1.

Plotting period (configuration variable `plotTS`)—time period for the simulated data to be displayed. It must be not smaller than the sampling period.

Timeout (variable `timeout`)—duration of a pause between plot updates. This parameter makes a difference only in graphical mode.

- Fundamental diagram editor (`Edit` → `Fundamental Diagram`, Figure 4.6). The triangular fundamental diagram in cell i is determined by either of two triplets: capacity (maximum flow) F_i, critical density ρ_i^c and jam density $\bar{\rho}_i$; or capacity F_i, free flow speed v_i and congestion wave speed w_i.

- Ramp data editor (`Edit` → `Ramp Data`) allows a user to modify on-ramp flows (Figure

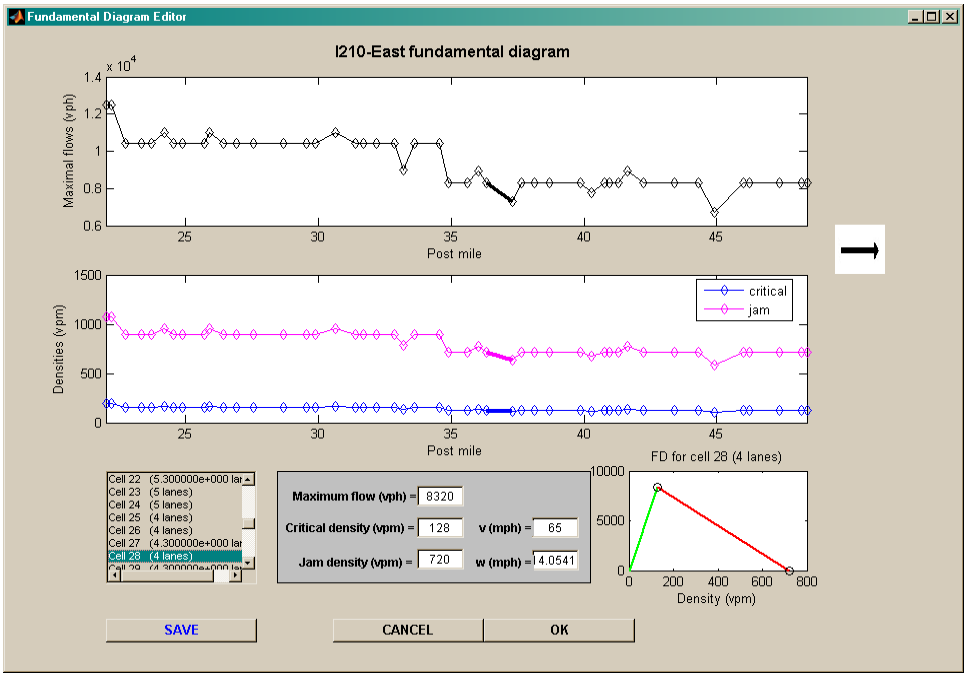


Figure 4.6: Fundamental diagram editor.

4.7a), or off-ramp split ratios (Figure 4.7b). All changes the user makes in the ramp data editor are overriden by demand or split ratio/off-ramp flow profiles if these are turned on.

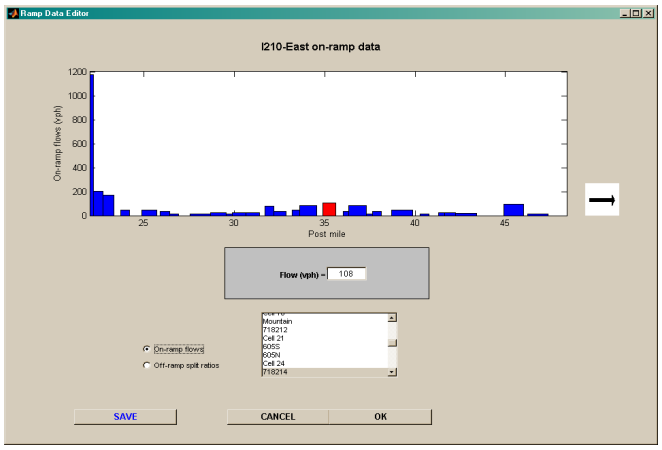
- On-ramp parameters editor (`Edit` → `On-Ramp` → `Parameters`, Figure 4.8a).

  Maximum flow—on-ramp capacity R_i.

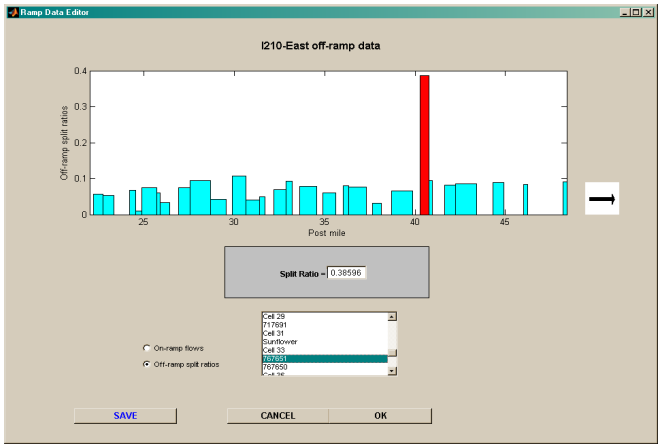
  Maximum queue size—number of vehicles the on-ramp can hold. This parameter is used only by queue controllers.

  Flow coefficient—serves to modify on-ramp demands coming from the profile without modifying profile itself.

  γ_i—on-ramp flow blending coefficient.



(a)



(b)

Figure 4.7: (a) Editor for on-ramp flows. (b) Editor for off-ramp split ratios.

ξ_i—on-ramp flow allocation coefficient.

- Off-ramp parameters editor (`Edit` → `Off-Ramp` → `Parameters`, Figure 4.8b).

  Maximum flow—off-ramp capacity S_i.

  Flow coefficient—serves to modify split ratios or off-ramp flows coming from the corresponding profile without modifying profile itself.

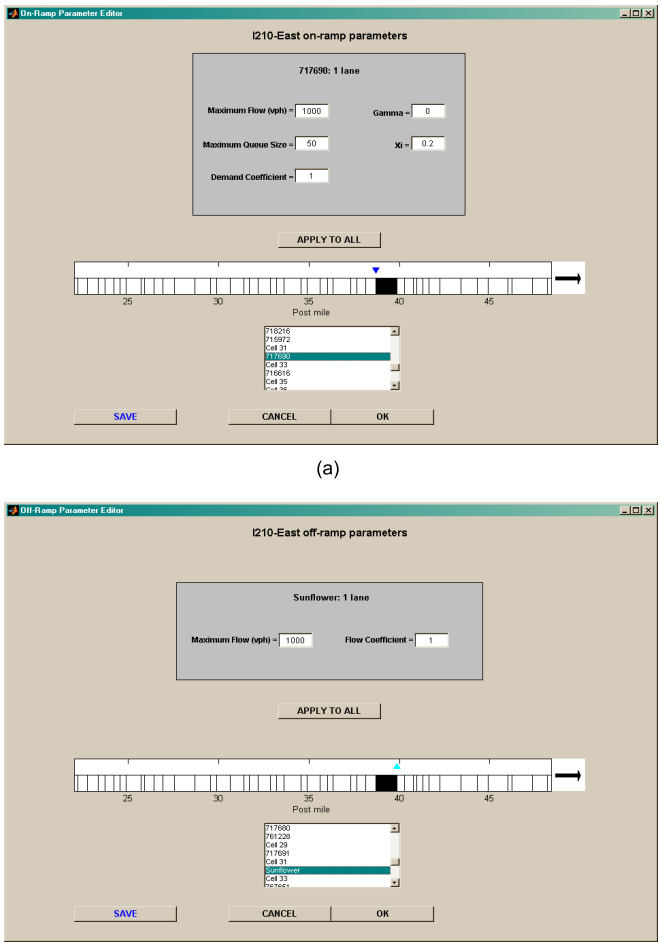


Figure 4.8: (a) Editor for on-ramp parameters. (b) Editor for off-ramp parameters.

Every on-ramp can have different mainline and queue controllers. Assigning particular controllers to on-ramps is done through the controller editor (Figure 4.9). This editor is also used to modify specific controller parameters. To call it, go to `Edit` → `On-Ramp` → `Controllers`. Section 4.2.3 explains how user can plug his/her own controllers into CTMSIM.

Each editor has `SAVE` button pressing which user can save his/her modifications in the current

configuration file.

Remark. If `SAVE` was pressed and the updates were saved, pressing `CANCEL` will not undo the changes to the configuration file, it only discards the changes for the current simulation. To undo the changes in the configuration file, press `CANCEL`, then call a particular editor again, then press `SAVE`.

To run `ctmsim` in batch mode, type

```
>> ctmsim myconfig 'b'
```

In this case, `ctmsim` saves the resulting simulation data in `.mat` file specified by the `dataFile` configuration variable.

4.2.3 Ramp Controllers

CTMSIM package contains file `ormlclist.mat` that carries the information about on-ramp mainline controllers. Currently implemented are ALINEA [90], LQI [98], SWARM [33] and the so-called *ideal ramp metering* (IRM) strategy.

The idea of IRM is simple. In given cell i with on-ramp, it selects the input flow r_i as

$$r_i = \max\left\{\min\left\{(\alpha\rho_i^c - \rho_i)\Delta x_i, d_i + q_i/(\Delta t), R_i\right\}, 0\right\}, \tag{4.18}$$

where coefficient α determines the desired portion of critical density below which we would like to keep the actual density. Default value is $\alpha = 0.97$. It means, the objective of the IRM controller is to keep the density in the current cell not higher than 97% of critical.

IRM strategy may not be so "ideal" in case there is an off-ramp in the current cell, or the capacity of the downstream cell is larger than that of the current one. In this situation, keeping the density in the current cell below critical may result in the underutilization of freeway capacity in the downstream cell. To avoid the underutilization, we implemented the *modified* version of IRM (MIRM) which, in addition to (4.18), considers the downstream capacity and adjusts input flow r_i accordingly: ρ_i^c in expression (4.18) is replaced by

$$\max\left\{\rho_i^c, g_i/v_i\right\},$$

with

$$g_i = \begin{cases} \bar{\beta}_i^{-1} \min\left\{\frac{\bar{\beta}_i}{\beta_i} S_i, F_{i+1}\right\}, & \text{if } \beta < 1, \\ S_i, & \text{otherwise.} \end{cases}$$

Adding new controllers to CTMSIM is fairly simple. Suppose, the user wants to implement his/her own controller *Xyz* and plug it into the CTMSIM. The following steps should be taken.

1. Perform clean up: delete all variables in the workspace.

   ```
   >> clear all
   ```

2. Load the list of on-ramp mainline controllers.

   ```
   >> load ormlclist
   ```

   Now, variable `mlclist` contains the list of already installed controllers. Other variables are specific controller structures.

3. Add Xyz controller to the list.

```
>> strvcat(mlclist, 'Xyz');
```

4. Define Xyz controller structure. Mandatory fields are listed in Appendix A.

```
>> xyz.id = size(mlclist, 1);
>> xyz.name = 'Xyz';
>> xyz.TS = 1/120;  % can be any other nonnegative value
>> xyz.Cmin = 0;  % can be any other nonnegative value
>> xyz.Cmax = 5000;  % can be any other value >= Cmin
>>
>> % other fields are user specific.
```

5. Save updated information in `ormlclist.mat` file.

```
>> save ormlclist
```

6. It remains to implement controller function for Xyz. It must satisfy the following format.

```
function new_orflows =
   controller_xyz(densities, orflows, celldata, ts, idx)
```

   where `densities` is a vector of densities, `orflows` is a vector of on-ramp flows, `celldata` is an array of cell data structures (see Appendix A for details), `ts` is a sampling period, and `idx` - cell index. The function should return the updated vector of on-ramp flows. See `controller_alinea.m` for an example of the ALINEA implementation.

After performing these steps, the Xyz controller is ready to use. In the `ctmsim` application, go to `Edit` → `On-Ramp` → `Controllers`. The controller editor window will pop up. Xyz controller is now in the list and it is possible to assign it to on-ramps (see Figure 4.9). All

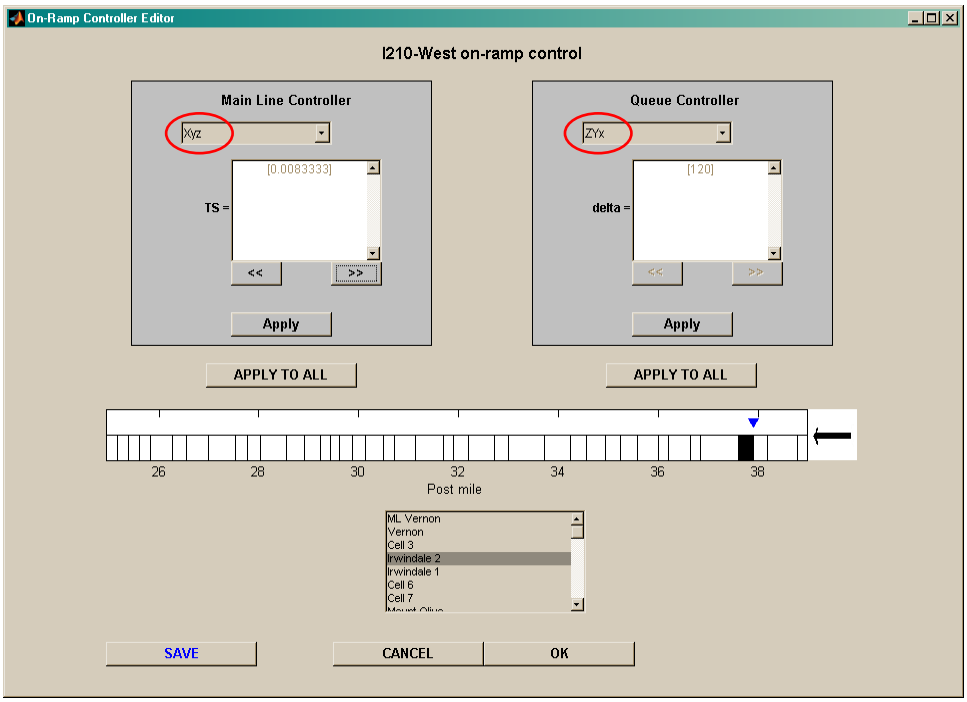


Figure 4.9: On-ramp controller editor window. Xyz mainline controller appears in the list. ZYx appears in the list of queue controllers.

controller parameters contained in the `xyz` structure can be adjusted in this window except `id` and `name`.

Similarly to `ormlclist.mat`, file `orqclist.mat` has information about queue controllers. Currently implemented queue control schemes are "queue-override" and proportional [98].

Adding user-designed *ZYx* queue controller to CTMSIM follows similar steps to those for mainline controller.

1. Perform clean up: delete all variables in the workspace.

   ```
   >> clear all
   ```

2. Load the list of on-ramp queue controllers.

```
>> load orqclist
```

Now, variable `qclist` contains the list of already installed queue controllers. Other variables are specific queue controller structures.

3. Add ZYx controller to the list.

```
>> strvcat(qclist, 'ZYx');
```

4. Define ZYx controller structure. Mandatory fields for queue controllers are also listed in Appendix A.

```
>> zyx.id = size(qclist, 1);
>> zyx.name = 'ZYx';
>>
>> % other fields are user specific.
```

5. Save updated information in `orqclist.mat` file.

```
>> save orqclist
```

6. It remains to implement the queue controller function for ZYx. It must satisfy the following format.

```
function new_orflows =
   controller_zyx(demands, orflows, orqueues, celldata, ts, idx)
```

where `demands` is a vector of on-ramp demands, `orflows` is a vector of on-ramp flows, `orqueues` is a vector of on-ramp queues, `celldata` is an array of cell data structures (see Appendix A for details), `ts` is a sampling period, and `idx` - cell index. The function should return the updated vector of on-ramp flows. See `controller_q_override.m` for an example of "queue override" implementation.

After these steps are performed, ZYx can be assigned to on-ramps through the controller editor (Figure 4.9).

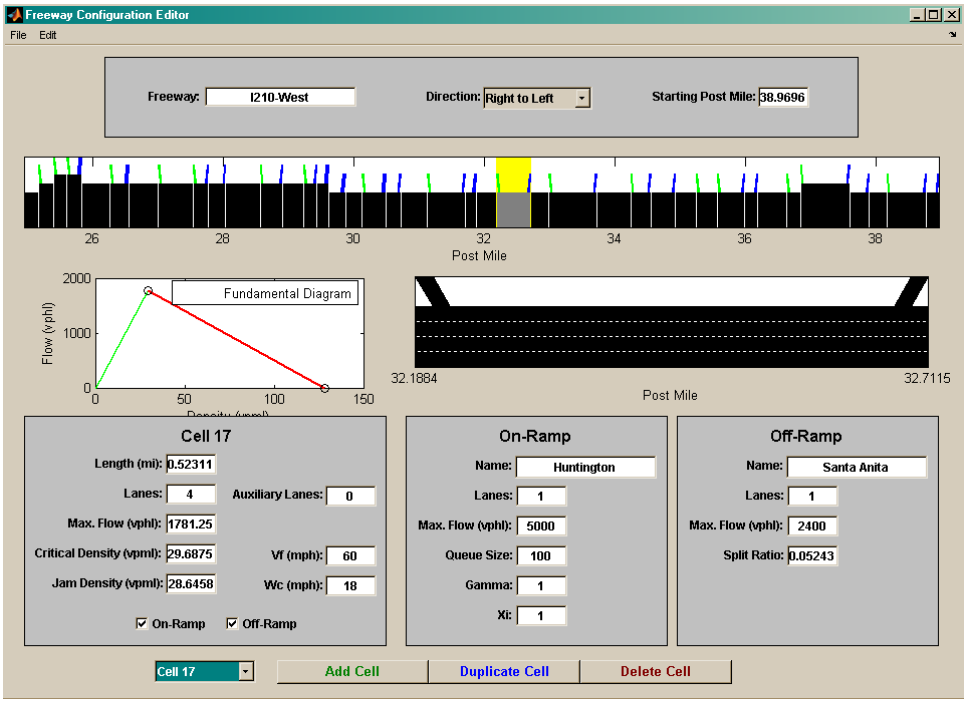


Figure 4.10: Freeway configuration editor `fwconfig`.

4.2.4 Auxiliary Utilities

Additionally to `ctmsim`, three other useful utilities are part of CTMSIM package. These are `fwconfig`, `plotsim` and `plotsim3`.

`fwconfig` is a graphical freeway configuration utility (see Figure 4.10). It allows users to build a freeway from scratch:

```
>> fwconfig
```

and save the result as CTMSIM configuration file; or edit an existing configuration:

```
>> fwconfig myconfig
```

where `myconfig.mat` is an existing CTMSIM configuration file.

More details about the format of configuration file can be found in Appendix A.

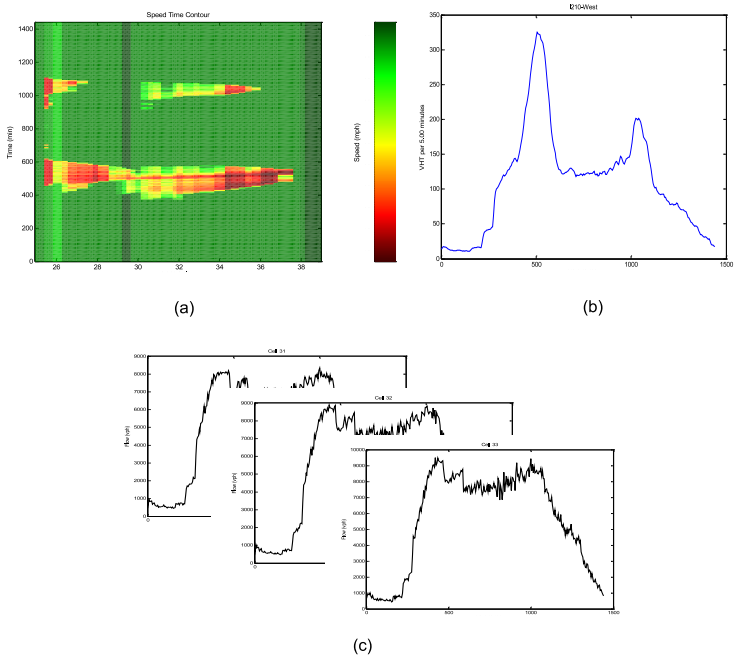


Figure 4.11: Using `plotsim` to plot (a) speed time contour; and (b) VHT evolution in time; (c) flow evolution in time in cells 31 through 33 in black color.

`plotsim` is a utility that plots requested simulation data in 2D:

```
>> plotsim mysim quantity
```

where `mysim.mat` is a file where CTMSIM simulation is saved, and the `quantity` parameter may assume one of the values listed in Table 4.2.

See Figure 4.11a,b for examples of `plotsim` command:

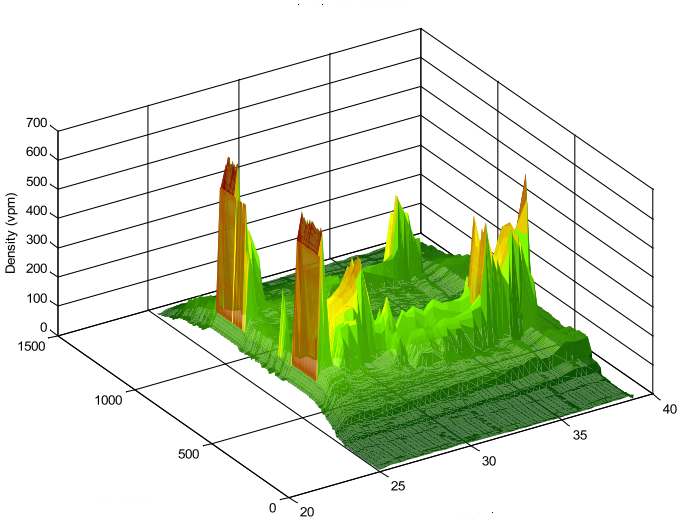


Figure 4.12: Using `plotsim3` to plot density in 3D. The axes are *Post mile*, *Time (min)* and *Density (vpm)*.

```
>> plotsim mysim 'speed'
>> plotsim mysim 'vht'
```

Additional parameters of command

```
>> plotsim('mysim', quantity, cells, color)
```

are `cells` - array of cell numbers for which the specified quantity should be plotted, and `color` - color for the plots. The result of the call

```
>> plotsim('mysim', 'flow', 31:33, 'k')
```

is shown in Figure 4.11c.

`plotsim3` plots simulated quantities listed in Table 4.2 except for travel time (`'ttime'`) in 3D. For example, the result of the call

Quantity	Description
`'density'`	Mainline density
`'flow'`	Mainline flow
`'speed'`	Traffic speed
`'ttime'`	Travel time
`'demand'`	On-ramp demands
`'orflow'`	On-ramp flows
`'queue'`	On-ramp queues
`'frflow'`	Off-ramp flows
`'beta'`	Off-ramp split ratios
`'vht'`	Vehicle Hours Traveled
`'vmt'`	Vehicle Miles Traveled
`'delay'`	Delay in vehicle hours
`'ploss'`	Productivity loss in lane mile hours

Table 4.2: Admissible values of `quantity` parameter.

```
>> plotsim3 mysim 'density'
```

is shown in Figure 4.12.

4.3 I210 Case Study

The first task undertaken by TOPl was a study of both directions of I210 freeway in Southern California. The segment of interest is beween junction with SR-134, postmile 25 (West bound) and Baseline Road, postmile 52 (East bound), see Figure 4.13.

To study each of the I210 directions, East and West, we went through the following steps.

1. We start by extracting the freeway geometry from PeMS: vehicle detector station (VDS) locations (post miles), number of lanes on each VDS-to-VDS segment, on-

Figure 4.13: Google map of I210.

and off-ramp locations (post miles), number of lanes at ramps. Currently, PeMS' information about freeway geometry is not complete. It only knows about ramps and lanes that have VDS. Therefore, after the freeway geometry is extracted from PeMS, we manually compare it with Google map and fill in missing lanes and ramps.

2. We divide the extracted freeway into cells in such a way that

   (a) any cell contains not more than one on-ramp, and if there is one, it is located at the beginning of the cell;

   (b) any cell contains not more than one off-ramp, and if there is one, it is located at the end of the cell;

   (c) any cell contains not more than one VDS;

   (d) cells are not too long—ideally, not longer than 1 mile - to make traffic dynamics more realistic;

   (e) cells are not too short—ideally, not shorter than .15 to .2 mile—to ensure that the sampling period Δt that has to satisfy (4.1) is not too small and, at the same

time, the number of cells not too large, thus making the simulation execute faster.

As Figure 4.14 shows, condition (c) is usually taken care of by the nature of VDS placements, which are rarely too close to each other. More often it is the case that some cells have no VDS at all. If two on-ramps are too close to each other, one may choose to treat them as one with demand equal to the sum of the two demands, to ensure (a) and (e). Similarly, collapsing two close off-ramps into one with double capacity, will ensure (b) and (e). If an on-ramp is closely followed by a off-ramp, the way to fulfill (a), (b) and (e) would be to place them in two separate neighboring cells: the off-ramp into the upstream and the on-ramp into the downstream cell.

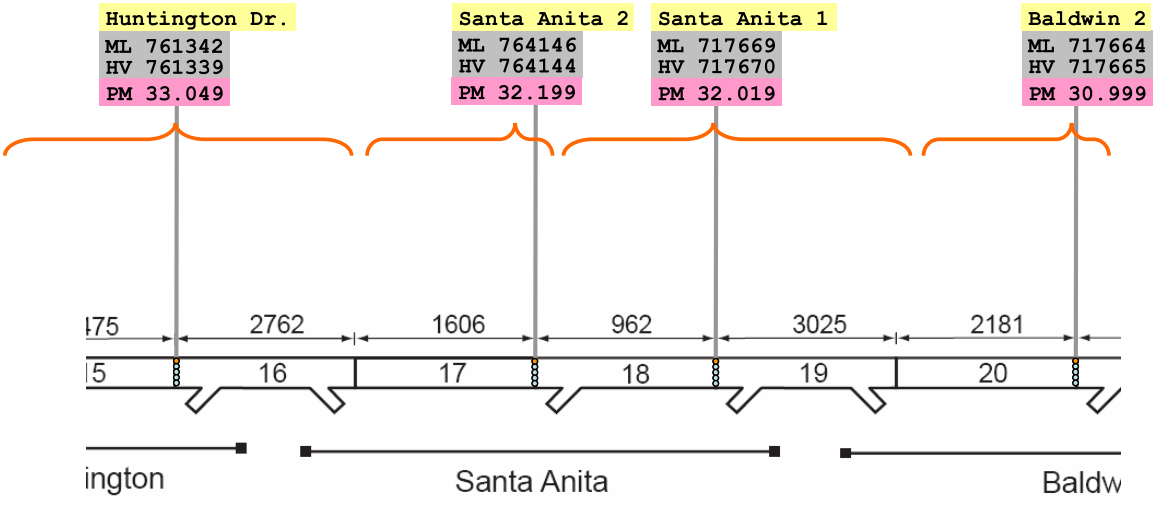


Figure 4.14: Segment of I210-West between Huntington Drive and Baldwin Avenue divided into cells, and correspondig VDSs. The cell length is in feet: cell 16 is 2762 feet long. Abbreviations: *ML* - mainline VDS, *HV* - HOV VDS, *PM* - post mile.

3. Once the freeway layout is in place, it is time to estimate the fundamental diagrams for each cell i, $i = 1..N$. This process is usually called *calibration* and is described in

[87, 100, 70]. In TOPl, we use none of these calibration methods in pure form. Rather, ours is a hybrid of several techniques and it is yet to be documented.

To summarize, calibration is performed by the following steps

(i) For each VDS in the freeway study segment extract density and flow values from PeMS for those days when this particular detector was in good working condition (PeMS also contains detector health status history).

Remark. The distance between neighboring VDSs is often larger than the cell size in CTM. Hence, the retrieved VDS data may apply (and usually does) to more than one cell (e.g. VDS 717669 covers cells 18 and 19 in Figure 4.14).

(ii) Find maximum flow value. Usually, this will be the capacity F_i.

(iii) Use least squares method to estimate free flow speed v_i. Practice shows that free flow density-flow pairs give a good fit.

(iv) Critical density $\rho_i^c = \frac{F_i}{v_i}$.

(v) Use constrained least squares method to determine congestion wave speed w_i.

When VDS data are good, steps (i-v) produce a decent result (Figure 4.15a). If, on the other hand, VDS data are poor due to malfunctioning detector or just because the capacity is never reached at this point of freeway as in Figure 4.15b, then we can either use fundamental diagrams from the neighboring cells, or impute the missing data [31] and repeat Step 3.

4. Determine "good days" for our freeway segment with respect to data collected by PeMS. "Good day" means a day when all or almost all vehicle detectors were healthy

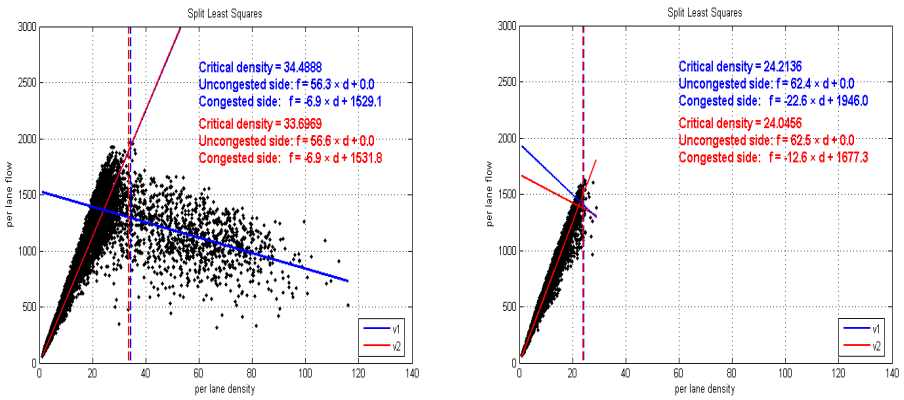


Figure 4.15: Estimating fundamental diagram. (a) Good data. (b) Poor data.

and collected data. In practice, it is never the case that all detectors are intact. Most off-ramp VDSs on I210 do not work. Thus, we relax the definition of a "good days" to those when mainline and on-ramp detectors were functional.

5. Create demand profiles for "good days" using PeMS flow data from on-ramp VDSs. Data for ramps without VDSs have to be imputed.

6. Compute off-ramp split ratio profiles with

$$\beta_i(t) = \frac{s_i(t)}{f_i(t) + s_i(t)}. \tag{4.19}$$

This is easier to say than to do. If VDS at an off-ramp doesn't work, we do not know $s_i(t)$. A rather rough but simple way around it, is by taking the flow data from upstream mainline and on-ramp VDSs (if these data are reliable, of course), estimate the sum $f_i(t) + s_i(t)$. Then, $f_i(t)$ can be obtained from the downstream VDS, and thus, we get $s_i(t)$.

7. Run the simulation with ramp control turned off for some of the "good day" demand and split ratio profiles and compare the results with PeMS data for those days. Because of the imputation of the missing on-ramp data and errors in estimating split ratios, adjustments to demand and split ratio profiles may be necessary. The interactive nature of `ctmsim` allows to tweak demand coefficients at particular on-ramps or modify split ratios as simulation executes. Figure 4.16 compares speed time contour from PeMS to one generated by `ctmsim`, and Figure 4.17 compares aggregate quantities: travel time, VHT, VMT and delay for a 14-mile segment of I210-West on a particular day, April 12, 2006.

   As simulation data matches well the actual freeway measurements from PeMS (Figures 4.16-4.17), we may conclude that CTM is adequate for traffic modeling, and demands and split ratios are adjusted well enough for given "good days".

8. It is time to start experimenting with on-ramp control strategies to see if they help improve, and how much, the freeway system performance. For example, Figure 4.18 shows how travel time and VHT under the no control strategy compares with those in the presence of ALINEA at on-ramps. In this example, although travel time along the freeway has reduced significantly by applying ALINEA, the overall performance has not improved much judging by the VHT chart. It means that while the freeway is in free flow almost all the time, large queues are formed at on-ramps.

   Note: VHT computed by PeMS does not include ramp delay, whereas they are included in VHT computed by CTMSIM.

9. While it may seem that certain control strategies offer little improvement (Figure

4.18b) under current conditions, it makes sense to check what will happen under different scenarios.

Scenario 1: 5% demand increase.

We can expect that two years from now the number of vehicles on the road will increase by 5%. In terms of our simulation settings, it means that the on-ramp demand values will increase by 5%. In the editor for on-ramp parameters, we set the value of demand coefficient to 1.05 for every on-ramp, and run the simulation with on-ramp control off and on.

Figure 4.19 shows that under increased demand, the impact of ALINEA is very significant. Ramp metering keeps the freeway in free flow, while delays imposed by queues at on-ramps are small compared to those resulting from congestion. Uncontrolled and ALINEA speed contours are presented in Figure 4.20.

Scenario 2: 2% demand decrease.

Now we check what happens if the number of vehicles on the road decreases. Such situation may be a result of a proper demand management (traveler information, tolls).

Say we want to decrease the demand by 2%. For that, we set the value of the demand coefficient to 0.98 in the editor for on-ramp parameters, and again run the simulation with on-ramp control off and on.

Figure 4.21 shows the travel time and VHT for the uncontrolled and ALINEA controlled cases, and Figure 4.22 shows the corresponding speed contour plots. As we can see, ramp metering does not improve the VHT performance. The delays are shifted

from the freeway to on-ramp queues. Ramp metering may still be beneficial, however, since it keeps the freeway in free flow maintaining a level of service that may be required by a tolling authority.

Scenario 3: accident.

Suppose, an accident occurred at 3.40 pm (minute 940 on the plot) at post mile 30, near Michllinda Avenue, leaving two of the four lanes blocked, and it took 20 minutes to clear the freeway. To simulate this accident, the user has to stop the running simulation at 940 minutes, open the fundamental diagram editor and modify the fundamental diagram of the cell where the accident took place (in our case, it is cell 25). Critical density, jam density and maximum flow for the selected cell should be reduced by 50%. Then the user has to resume the simulation, stop it at 960 minutes, change the fundamental diagram for that cell back to the original, resume the simulation again and run it to the end. This simulation has to be run with on-ramp control off and on.

Figure 4.23 shows the travel time and VHT for the uncontrolled and ALINEA controlled cases, and Figure 4.24 shows the corresponding speed contour plots. Clearly, ramp metering significantly improves both travel time and VHT at the time of accident and after and keeps the freeway mostly in free flow.

The user may choose to simulate other variations of this scenario, such as increasing or reducing the time interval between the accident occurrence and clearance.

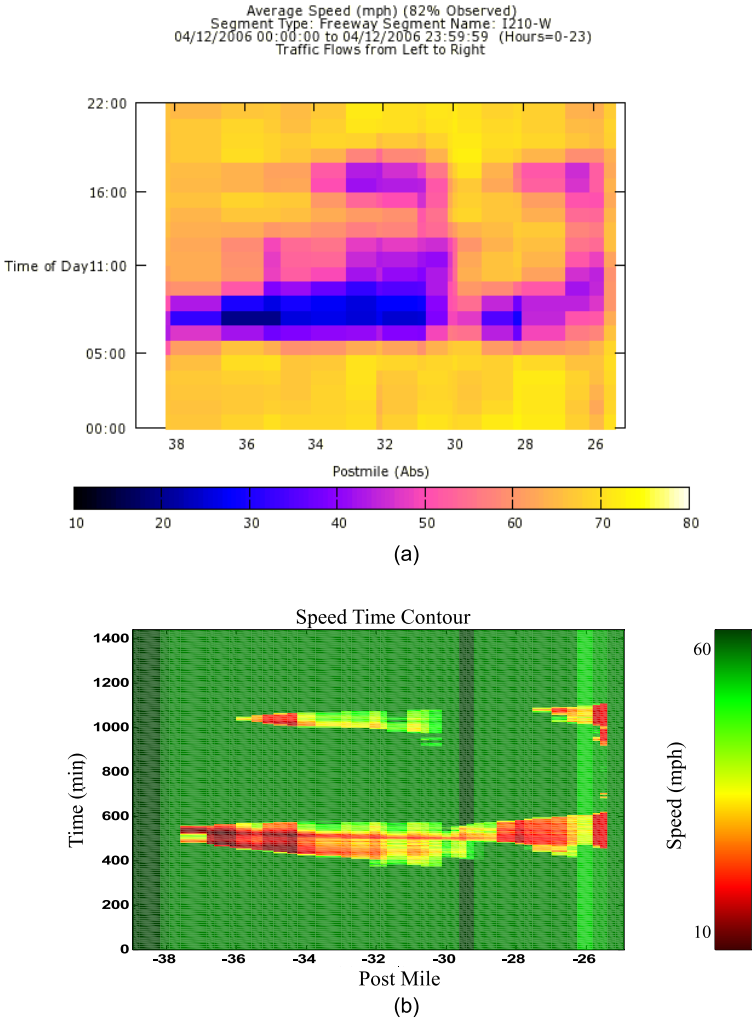


Figure 4.16: I210-West 14-mile segment beween Vernon Avenue and junction with SR-134, April 12, 2006. Speed time contours (traffic flows from left to right).
(a) PeMS data.
(b) CTMSIM simulation.

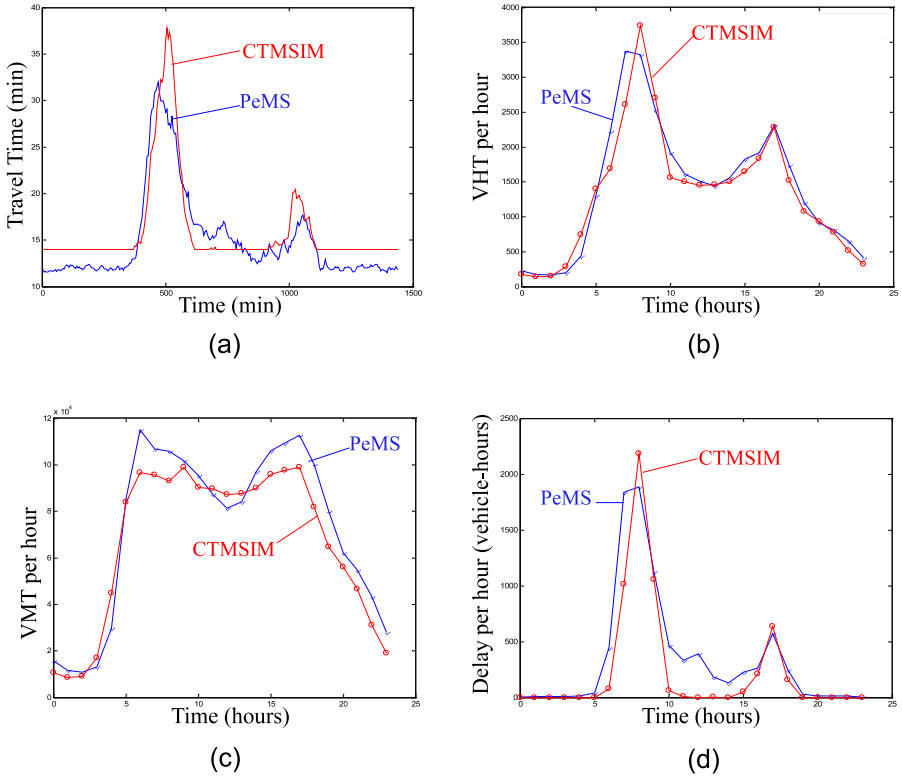


Figure 4.17: I210-West 14-mile segment beween Vernon Avenue and junction with 134, April 12, 2006. Aggregate data comparison between PeMS (blue) and CTMSIM (red).
(a) Travel time. (b) VHT. (c) VMT. (d) Delay.

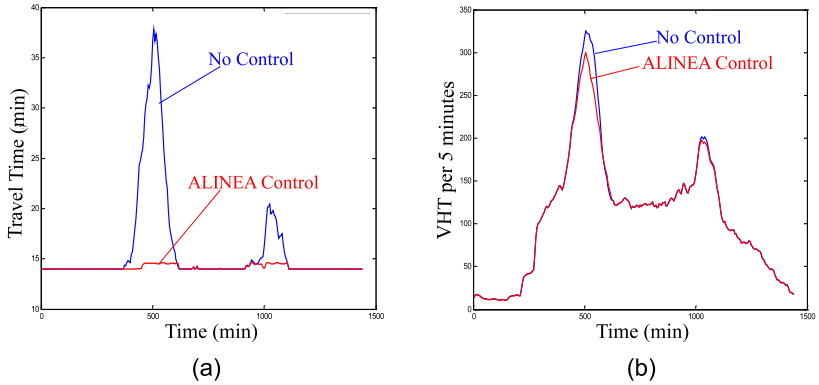


Figure 4.18: Freeway performance without control (blue) vs. with ALINEA control (red).
(a) Mainline travel time. (b) VHT.

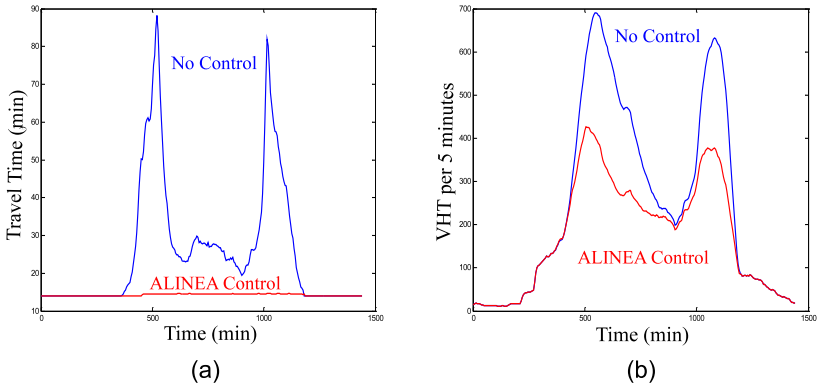


Figure 4.19: Scenario 1: 5% demand increase. Freeway performance without control (blue) vs. with ALINEA control (red). (a) Travel time. (b) VHT.

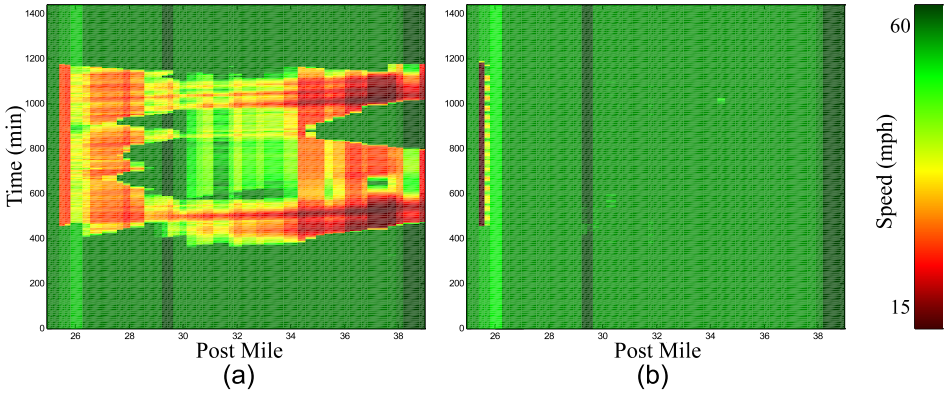


Figure 4.20: Scenario 1: 5% demand increase. Comparison of two speed time contours (traffic flows from right to left). (a) No control. (b) ALINEA control.

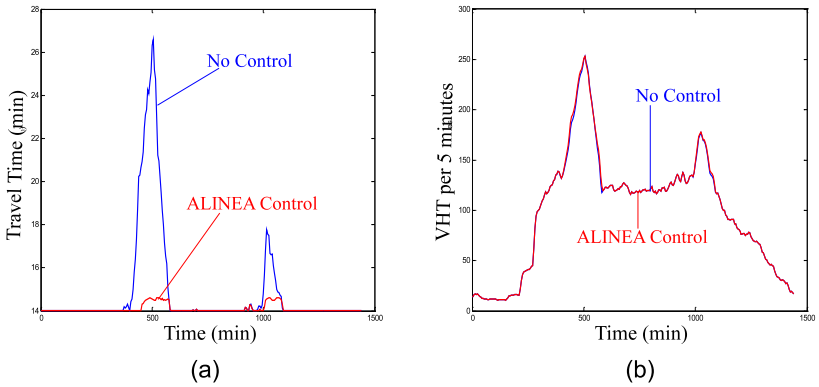


Figure 4.21: Scenario 2: 2% demand decrease. Freeway performance without control (blue) vs. with ALINEA control (red). (a) Travel time. (b) VHT.

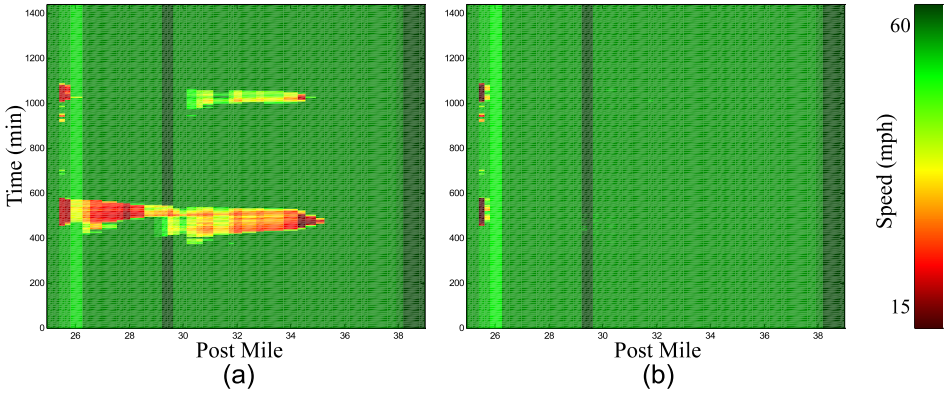


Figure 4.22: Scenario 2: 2% demand decrease. Comparison of two speed time contours (traffic flows from right to left). (a) No control. (b) ALINEA control.

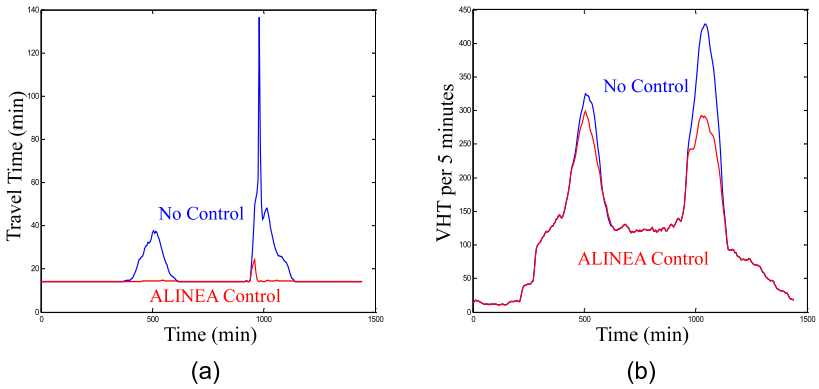


Figure 4.23: Scenario 3: accident at post mile 30 (near Michillinda Avenue). Freeway performance without control (blue) vs. with ALINEA control (red). (a) Travel time. (b) VHT.

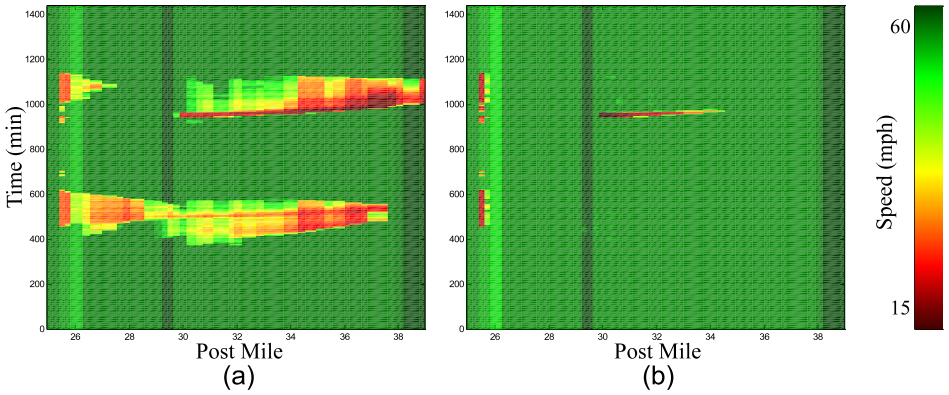


Figure 4.24: Scenario 3: accident at post mile 30 (near Michillinda Avenue). Comparison of two speed time contours (traffic flows from right to left). (a) No control. (b) ALINEA control.

Chapter 5

Aurora: Simulation and Analysis Framework for Infrastructure Networks

5.1 Motivation

Although it is a neat software package, useful for transportation researchers in their studies of freeway traffic, CTMSIM falls short of providing an appropriate toolset for corridor management. It handles only one freeway and no arterial networks, cannot deal with HOV, has no notion of *event* that triggers certain configuration or input changes at given time or under given conditions allowing to program scenarios, and requires MATLAB which makes it unusable for the operations staff in organizations such as Caltrans.

The TOPl group had an option: to find a third party macro-simulation package able to handle mixed freeway-arterial road network and adapt it to its needs, or develop its own software solution. Of the third party packages, METACOR [44] suited best the needs of TOPl, but its commercial version [5] was unavailable to us at the time (December 2006), and it was decided to proceed with development of our own simulation and analysis framework for traffic flows in freeway corridors. Aurora[1] [9] is the working name given to this project.

Listed below are the foundation principles of the Aurora framework.

1. *Multi-purpose*—basic structures and algorithms must be generic and not road traffic specific, making the framework reusable for other applications, such as irrigation canals, oil or gas pipelines, etc. Application specific classes inherit from these basic structures. This affects basic data structure definitions and general purpose algorithm development.

2. *Usability*—Aurora tools must be easy to handle: creating configuration files, running simulations, and calling analysis routines, must be intuitively clear. It is better to have several different lightweight applications for different tasks rather than one heavyweight application for multiple purposes.

3. *Interactivity*—simulation and analysis applications must provide clear and simple GUI with good data visualization. This and previous items affect the user interface and visualization: what data should be displayed and in what way.

[1] Aurora (short for polar aurora) - glow in the sky, seen often in a ring-shaped region around the magnetic poles ("auroral zone") and occasionally further toward the equator. The name comes from an older one, “aurora borealis”, Latin for “northern dawn”, given because an aurora near the northern horizon (its usual location when seen in most of Europe) looks like the glow of the sky preceding sunrise. Also known as “northern lights”, although it occurs both north and south of the equator.

4. *Scenario oriented*—user should be able to write scenarios: lists of events that change configuration or inputs, with times or conditions of their occurrence, and feed them to the simulator. This affects the way events are described and handled.

5. *Scalability*—user should be able to seemlessly add new roads (canals, power lines, etc.) to already existing network configurations, or connect two or more networks with each other. This affects basic data structure definitions and the way configuration files are organized.

Aurora is implemented in Java. It uses external libraries: JUNG [10] for drawing the network, and JFreeChart [11] for plotting simulation data.

Java packaging of Aurora is organized as follows:

- *aurora*—generic classes and interfaces;
- *aurora.hwc*—classes and interfaces specific to road network application (HWC stands for highway control);
- *aurora.hwc.gui*—user interface classes for the road network application;
- *aurora.util*—collection of useful routines, not application specific.

5.2 Architecture

5.2.1 Basic Objects

The basic building block of the Aurora system is a *network element* with a unique integer ID. A network element can be a *link* representing a stretch of road (water canal, pipeline, etc.), *simple node*—point where links merge and/or diverge, *complex node*—network built out of network elements, or *monitor*—an object that monitors the state of specified links and nodes in a network.

A link has direction and length. It must have either of the two nodes, *begin node* or *end node*, or both of them, attached to it. Links with no begin nodes are *source links*. Source links provide input to a system. In the case of a road network (Aurora HWC[2]), associated with source links are demand values and queues. Links with no end nodes are *destination links*. In Aurora HWC, we assume that anything downstream of a destination link is in free flow.

Table 5.1 summarizes link types implemented in Aurora HWC together with begin and end nodes that each link type allows. Each of these link objects has associated with it a fundamental diagram and dynamics. Dynamics is an interface, i.e., any macroscopic traffic model can be used to compute the link state, namely, traffic density. Currently, only the CTM model is implemented. Density is implemented not as a simple numeric type, but as generic object, allowing us to modify or extend the traffic model so it could deal with density

[2]Aurora for highway control

Link type	Admissible begin nodes	Admissible end nodes
freeway	freeway	freeway
highway	highway	highway
HOV	freeway, highway	freeway, highway
interconnect	freeway, highway	freeway, highway
on-ramp	signal and stop junctions	freeway, highway
off-ramp	freeway, highway	signal and stop junctions
street	signal and stop junctions	signal and stop junctions
dummy	any	any

Table 5.1: Aurora HWC link types with corresponding admissible begin and end nodes.

as a vector of values (say, when traffic flows carry additional origin-destination information, or we want to distinguish vehicles by types, e.g. SOV, HOV, trucks), or more complex data structure. For each link, Aurora computes travel time, VHT, VMT, delay and productivity loss.

Node type	Admissible input links	Admissible output links
freeway	freeway, HOV, interconnect, off-ramp	freeway, HOV, on-ramp, interconnect
highway	highway, HOV, interconnect, off-ramp	highway, HOV, on-ramp, interconnect
signal junction	street, off-ramp	street, on-ramp
stop junction	street, off-ramp	street, on-ramp

Table 5.2: Aurora HWC node types with corresponding admissible input and output links.

A simple node[3] must have one or more input and one or more output links. Aurora HWC nodes are listed in Table 5.2 together with types of input and output links they admit. Local controllers (such as ALINEA [90]), if any, reside on nodes and are assigned to given input links, potentially restricting flows coming from these links. When there are multiple output links, nodes also carry information about what portions of which input flows must

[3]We refer to it as node from now on, while referring to a complex node as a network.

be directed to which output. Currently, for m inputs and n outputs in the node, it is implemented as an $m \times n$ split ratio matrix, where elements are nonnegative and sum up to 1 in each row.

Remark. Currently, we do not distinguish between freeways and highways. Highway specific objects are present in Aurora following the road classification provided by HCM [101] and are reserved for future use.

While links and nodes physically form a network, a monitor is a special object whose purpose is to monitor the state of specified links and nodes, and based on these data, issue instructions to system wide controllers (such as SWARM [33]). Monitors are ad-hoc objects designed by the user together with network wide control strategies. They can be also used to generate certain events (such as split matrix change—to simulate traveler information affecting traffic flow directions) based on observed conditions (more about events in the next Section). At this point, no monitors are implemented.

All described network elements—links, nodes and monitors—are always part of a complex node, a network. There is at least one network in any Aurora system—the top level complex node, to which all other links, nodes and monitors belong. Network objects are nodes, hence, networks can contain networks just as they contain simple nodes. It makes the Aurora structure hierarchical, allowing to create configurations out of building blocks that are more complex than links and simple nodes, which is faster and more convenient, and opens a door to parallel computation when simulation steps for different subnetworks can be computed concurrently by different processors. Another benefit of using a hierarchical structure is that different subnetworks may have different sampling periods, that is, simulation steps

of different duration. It can save time if, for example, network consists of roads with long enough links that do not require a small sampling period, and roads with rather short links that do (recall (4.1)). Separating them into subnetworks with different sampling periods reduces computation time.

Remark. Sampling periods of subnetworks cannot be greater than sampling period of top level network.

There are two other basic objects. Object *path* describes route from a node to node as a sequence of adjacent links. For each path, Aurora computes travel time, VHT, VMT, delay and productivity loss based on corresponding data from links forming the path. Object *OD* describes a pair of origin and destination nodes together with list of paths connecting the two. A complex node may contain a list of origin-destination pairs. For consistency, it is required that every link in every path of every origin-destination pair belongs to the same complex node as ODs in the list[4].

5.2.2 Events

Aurora HWC-specific events, summarized in Table 5.3, are derived from the generic Aurora event object and are handled by the Aurora event manager. Events that change fundamental diagrams can be used to simulate traffic incidents by reducing capacity. Changing demand coefficients and split ratio matrices help imitate special events, road closures, or effects of

[4]It may happen that both, origin and destination nodes, belong to the same subnetwork, while some links at certain paths connecting them are part of a different subnetwork. This is not a problem because top level network contains all the links present.

displaying traveler information. Controller and queue size changes may be part of complex ramp control strategies.

Event: change in	Where occurs
controller	nodes
split ratio matrix	nodes
demand coefficient	source links
queue size	source links
fundamental diagram	links

Table 5.3: Aurora HWC events.

The event object carries the following information: activation time (in terms of simulation hours), ID of a network element where it must occur, new parameter values for this network element, and an `activate()` method that changes those parameter values at a network element while storing the old values coming from the network element in their place. An event list is an optional part of the configuration file, but the user can generate new events before or during the simulation run as well as disable those already in the list.

When the simulator reads a configuration file, it places all events listed there into a queue sorted by event activation time (this queue may be empty, if no events are specified). User generated events are added to this queue, their location in the queue being determined by their activation time also. Events are triggered by the event manager. Before each simulation step, it selects those events in the queue that are due (those whose activation time is smaller than the next simulation time step), and activates them by invoking their `activate()` methods. Then, it moves these activated events from the event queue to the end of the event history list in the order they occured. Thus, events never get deleted. Later, when the user resets the simulation, events are rolled back, or activated in reverse sequence

with reverse action, returning to network element parameters their original values. Such maintenance of event queue and history list potentially allowes us to "rewind" simulation to any given point of its execution.

5.2.3 Simulation Algorithm

An object representing a network element contains the `dataUpdate()` method. It performs simulation step computations specific to the particular type of network element. The recursive algorithm of `dataUpdate()` in a network is described next.

1. Check if at this time step any action is needed:

$$(k - k_0)\Delta t_0 < \Delta t, \tag{5.1}$$

   where Δt is sampling period for this network, Δt_0 is the sampling period for the top level network, k is the current time step, k_0 is the time step at which the last action was performed.

   If $k > 1$ and inequality (5.1) holds, then return without doing anything. Else, proceed to step 2.

2. For every monitor in the monitor list, call `dataUpdate()`. If present, each monitor has its own specific task—it may assign controller parameters, or generate events to be activated before the next simulation step or later, at prescribed time.

3. For every node in the node list, call `dataUpdate()`. If the node is complex, start the algoritm from step 1 with respect to this node. Else (the node is simple), compute

input and output flows based on demand from upstream and available capacity of downstream links. This can be done in many ways.

Daganzo in [40] introduces the concept of priorities for multiple input flows and the FIFO[5] rule for multiple output flows.

In the Aurora framework priorities are the fractions of the input flows accepted by the node, in case the upstream demand exceeds the downstream capacity (if the upstream demand is below the downstream capacity, priorities do not matter since all the vehicles from the upstream links can be accommodated by the downstream links). Different priority choices result in different flow values for the next simulation step. In the current Aurora implementation we assume that the input priorities are proportionate to the input demands.

The FIFO rule means that if one of the output links cannot accommodate its allocation of flow, the total output flow is restricted[6]. In the Aurora framework the FIFO rule implies that the input-output flow relations defined by the split ratio matrix must be preserved.

To summarize, we compute the input and output flows based on the input demands, satisfying the downstream capacity restrictions by assuming the input priorities to be proportionate to the demands, while preserving the input-output flow relations defined by the split ratio matrix.

Given $m > 0$ input and $n > 0$ output links, computation proceeds as follows:

[5]First in, first out.

[6]Vehicles unable to exit from the upstream link prevent all those behind, regardless of their destination, to continue.

(a) Compute input demands

$$\tilde{d}_i(k) = \min\left(v_i\rho_i(k_0), \mathcal{C}(\rho_i(k_0)), F_i\right), \quad i = 1..m, \tag{5.2}$$

where v_i is free flow speed, $\rho_i(k_0)$ is the density at the input link i; $\mathcal{C}(\rho_i(k_0))$ denotes flow value suggested by a controller, if a controller is assigned to the input link i; and F_i is the capacity of the input link i.

(b) From the $m \times n$ split ratio matrix $\mathcal{B}$, and the m-dimensional demand vector $\tilde{d}(k)$ we get the input-output $m \times n$ demand matrix $D(k)$,

$$D_{ij}(k) = \mathcal{B}_{ij}\tilde{d}_i(k), \quad i = 1..m,\ j = 1..n, \tag{5.3}$$

and output demands

$$d_j(k) = \sum_{i=1}^{m} D_{ij}(k), \quad j = 1..n. \tag{5.4}$$

(c) Compute available output capacities

$$c_j(k) = \min\left(w_j(\bar{\rho}_j - \rho_j(k_0)), F_j\right), \quad j = 1..n, \tag{5.5}$$

where w_j is congestion wave speed, $\bar{\rho}_j$ is the jam density, and F_j is the capacity of the output link j.

(d) Compute input-output demand matrix adjusted by the output link capacity restrictions, assuming the input priorities to be proportionate to the demands,

$$\hat{D}_{ij}(k) = \frac{\min(d_j(k), c_j(k))}{d_j(k)} D_{ij}(k), \quad i = 1..m,\ j = 1..n, \tag{5.6}$$

and adjusted input demands

$$\hat{d}_i(k) = \sum_{j=1}^{n} \hat{D}_{ij}(k), \quad i = 1..m. \tag{5.7}$$

This step ensures that the adjusted input demand does not exceed the downstream capacity. More precisely,

$$\sum_{i=1}^{m} \hat{D}_{ij}(k) \leq c_j(k),$$

with equality being achieved if and only if $d_j(k) \geq c_j(k)$.

Remark. Expression (5.6) makes sense only if $d_j(k) \neq 0$. So, in case $d_j(k) = 0$, we set $\hat{D}_{ij}(k) = 0$.

(e) Compute input flows

$$\tilde{f}_i = \hat{d}_i \min_j \left\{ \frac{\hat{D}_{ij}}{\hat{d}_i \mathcal{B}_{ij}} \right\}, \quad i = 1..m,\ j = 1..n. \tag{5.8}$$

In case $\hat{d}_i = 0$ or $\mathcal{B}_{ij} = 0$ for all $j = 1..n$, we set $\tilde{f}_i = 0$.

(f) Compute output flows

$$f_j(k) = \sum_{i=1}^{m} \mathcal{B}_{ij} \tilde{f}_i, \quad j = 1..n. \tag{5.9}$$

Steps (e) and (f) implement the FIFO rule: input and output flow values are assigned so as to maintain input-output relationship defined by matrix $\mathcal{B}$.

4. For every link in the link list, call `dataUpdate()`.

(a) Compute density and speed using model specific equations. For CTM, these are

$$\rho(k) = \rho(k_0) + \frac{\Delta t}{\Delta x}(f_u(k) - f_d(k)), \text{ and } V(k) = f_u(k)/\rho(k), \tag{5.10}$$

where Δx is the link length, f_u is the upstream flow (flow entering the link), f_d is the downstream flow (flow exiting the link), and V is the speed. If the link is

a source link, $f_u(k)$ equals current demand, otherwise $f_u(k)$ is computed in the begin node of the link as one of its output flows. If the link is a destination link,

$$f_d(k) = \min\left(v\rho(k_0), F\right),$$

where v is the free flow speed, and F is the capacity; otherwise $f_d(k)$ is computed in the end node of the link as one of its input flows.

(b) Compute travel time

$$TT(k) = \Delta x / V(k). \tag{5.11}$$

(c) Compute VHT, VMT, delay and productivity loss as described in Section 4.2, by formulas (4.10), (4.12), (4.14), (4.16).

5. Set $k_0 = k$ and return.

5.2.4 Configuration

From TOPl we learned that once the process of a freeway corridor study is established, the most tedious and time consuming task is putting together a configuration file with road network description. Being the least rewarding, this task requires attention to details and patience. Therefore, efficient configuration management was made one of the priorities of TOPl.

General configuration file contains

- information about network layout: positions and types of nodes, shapes and types of links connecting them, numbers of lanes;

- description of monitors (if any are present);
- split ratio matrices and controller information at nodes and fundamental diagrams at links;
- list of origin-destination pairs, each with list of feasible paths;
- demand profile for source links;
- event scenario - list of events describing what occurs, where and when.

The first item is the most difficult, because there is no single source from which these data could be extracted. Eventually, PeMS intends to provide this information for arterials as well as for freeways. To date, however, PeMS only deals with freeways and it knows only about ramps and lanes where detectors are installed. Thus, we have to work with GIS data from regional planning agencies such as MTC (Bay Area) and SANDAG (San Diego), configuration files for different simulators used by other research groups, and ultimatly consult Google maps. The procedure of network layout extraction is not well defined yet and requires a "human touch", i.e., a lot of manual checks and adjustments.

Calibration, i.e., computation of fundamental diagrams and split ratio matrices, and demand generation, with the lack of sufficient measurement data (especially, for arterials), is the second great challenge.

All this put together makes us realize that complete configuration files have significant value, so that establishing a repository of configurations makes sense.

Aurora configuration files use XML format whose syntax, data sets and validation rules, are defined by a schema presented in Appendix B. Types of links, nodes, monitors, controllers and dynamics are defined in `class` attribute, which specifies what classes Aurora must instantiate upon reading the configuration. Configuration is modular. That is, origin-destination lists, demand profiles and event scenarios, are separate blocks that can be optionally added to a configuration file or stored on their own. This, plus the hierarchical structure of Aurora in which a network is just another complex node, make the manipulation of configuration building blocks relatively easy and efficient.

Another benefit of XML configuration is that it can be read by anyone (the technology is known and proven) and translated into other formats. That makes it a good candidate for an interchange format for road network descriptions.

5.3 User Interface

Figure 5.1 presents the face of the Aurora HWC simulator. The application window is divided into four frames.

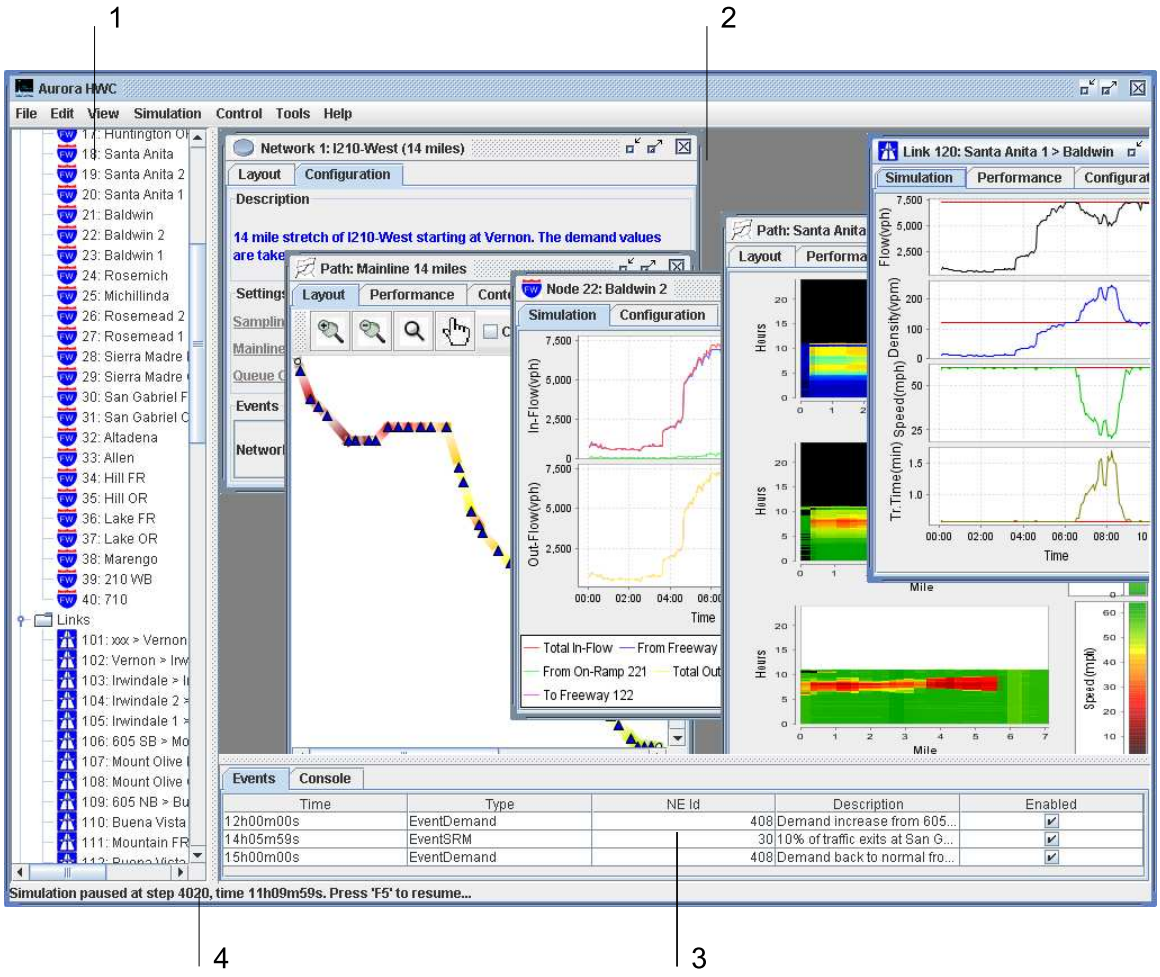


Figure 5.1: Aurora HWC simulator window.
1 - network tree frame: lists networks, nodes, links, monitors, ODs, paths.
2 - main frame: home to application subwindows.
3 - scenario frame: lists events and logs.
4 - status frame: displays simulation status, and/or instructions to the user.

1. Network tree frame. It displays network components - nodes, links, monitors, origin-destination pairs and paths in a hierarchical tree structure a la Windows Explorer. Special icons specify their types. Double clicking on a component brings up a subwindow in the main frame with details of that particular network element or path.

2. Main frame. It is used to host subwindows for selected network elements or paths.

    - Network subwindow (Figure 5.2). It is divided into two tabs. One displays network layout that can be zoomed in and out and allows to view details of links and nodes composing the network. As simulation runs, links are colored using either their density or speed values (user choice). The other tab describes the network together with general settings: sampling period, mainline and queue control on or off.

    - Node subwindow (Figure 5.3). It is divided into two tabs. One displays simulation data: input and output flows. The other displays node configuration: input links with controllers, output links and split ratios. From the configuration tab the user can generate node events (controller and split ratio matrix changes).

    - Link subwindow (Figure 5.4). It is divided into three tabs. Simulation tab displays flow, density, speed and travel time. Performance tab shows VMT, VHT, delay and productivity loss at the link. Configuration tab has information about link length, number of lanes and fundamental diagram. From here the user can generate link events (fundamental diagram changes, and demand and queue limit changes for source links).

    - Path subwindow (Figure 5.5). It is divided into three tabs. Layout tab shows

Figure 5.2: Network subwindow. (a) Layout tab. (b) Configuration tab.

part of network that constitutes a path. It can be zoomed in and out and allows to view details of participating links and nodes. As the simulation runs, links are colored using either their density or speed values (user choice). Performance tab shows travel time, VMT, VHT, delay and productivity loss along the path. The time granularity of the plotted VMT, VHT, delay and productivity loss values is determined by the display update period, which is part of the general settings. Contour tab displays flow, density and speed time contours for the path.

3. Scenario frame (Figure 5.1). This frame is divided into two tabs. One lists the events

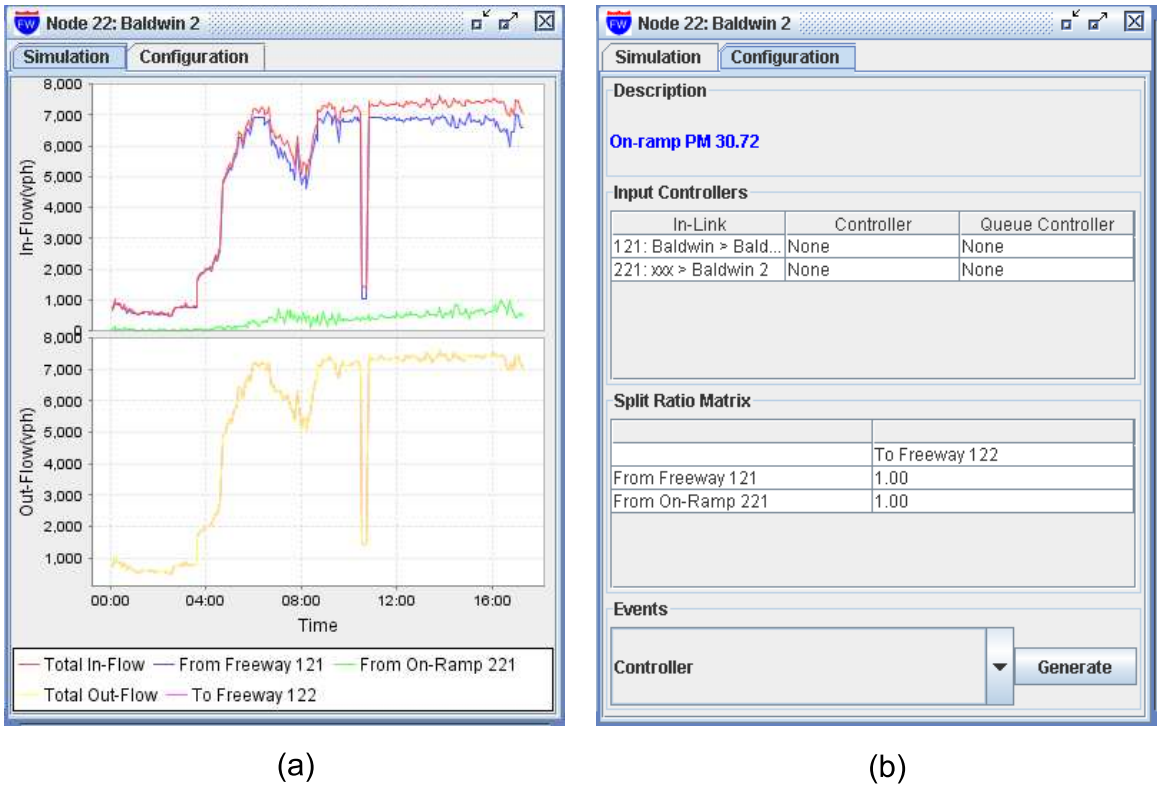


Figure 5.3: Node subwindow. (a) Simulation tab. (b) Configuration tab.

of the current simulation scenario: their type, description, activation time. The user cannot delete an event from the list, only edit it, or disable it. The event editor window pops up when the user double clicks on an event. Figure 5.6 shows an editor for the fundamental diagram change event. Each event type has its own editor window. The other tab is a console for dumping the application output: logs and debugging information.

4. Status frame (Figure 5.1). It displays the status of simulation—running, paused or stopped, and issues short instructions to a user.

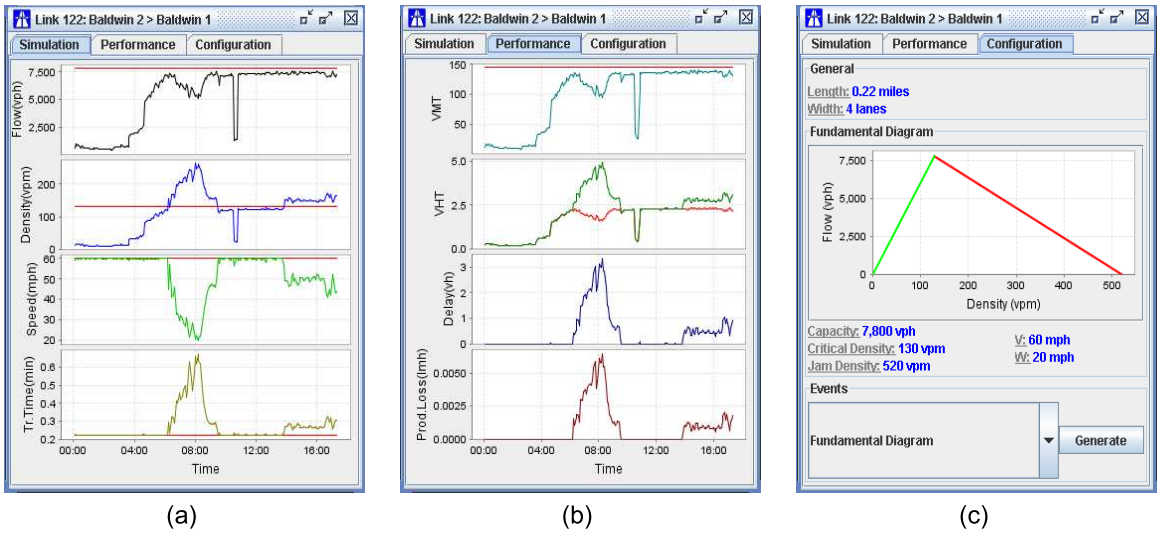


Figure 5.4: Link subwindow. (a) Simulation tab. (b) Performance tab. (c) Configuration tab.

The menu on top of the main window (Figure 5.1) currently provides the following options:

- to run simulation (`Simulation` → `Run`);
- to stop simulation (`Simulation` → `Stop`);
- to save simulation (`File` → `Save Simulation`);
- to reset simulation (`File` → `New Simulation`);
- to load previously saved simulation (`File` → `Open`);
- to toggle mainline and queue control switches (`Control` → `Mainline` and `Control` → `Queue`); to turn mainline control on and off;
- and to edit general settings (`Edit` → `Settings`, Figure 5.7).

  General settings comprise such parameters as display update period, maximum simulation time and timeout between screen refreshes.

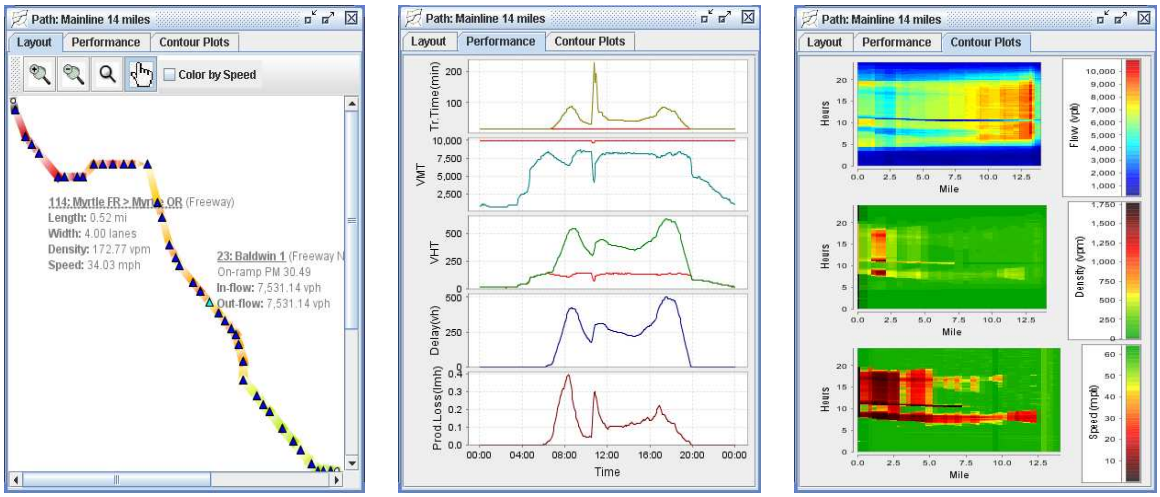


Figure 5.5: Path subwindow. (a) Layout tab. (b) Performance tab. (c) Contour tab.

It is also possible to save the current configuration (`File` → `Save Configuration`). Although a user has no direct way of changing configuration parameters of network elements in the simulator, they may change nevertheless, due to activated events. Also, a user can generate new, edit or disable old events, thus modifying event scenario. These are the reasons why it may be desirable to save current configuration in a file.

5.4 Goals

Aurora is not a finished product as of the moment this dissertation is written. As was mentioned above, it consists of three major modules: configuration, simulation and analysis.

Simulation module.

Simulation module is a centerpiece of the framework. Significant progress has been made in

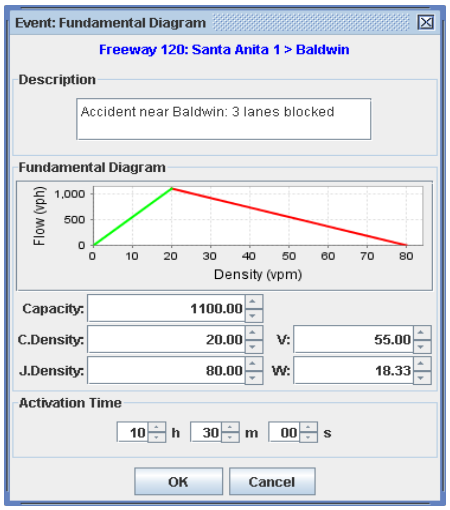


Figure 5.6: Editor window for the link event that changes fundamental diagram.

the simulator development, but several items remain unfinished.

- Modeling of arterial traffic requires implementation of signal control for nodes representing signal junctions. Until it is implemented, we make trivial assumption that arterial traffic always moves with free flow speed ($25-30$ mph) and there are no delays at signal junctions.

- Currently, we use demand values at source links as inputs to the system, while split ratio matrices at MIMO[7] nodes determine how traffic flow is divided between different links.

  Alternative form of input data can be used: origin-destination flow matrices. That is, instead of demand profile and split ratios at junctions, there are OD matrices generated with given period (say, every 5 or 15 minutes) that specify how many vehicles started

[7]Multiple input, multiple output.

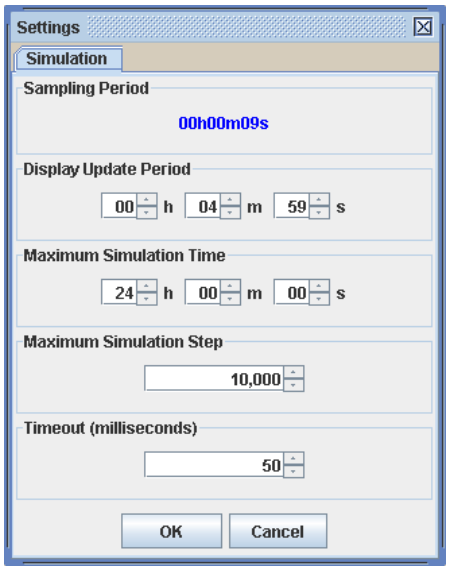


Figure 5.7: Editor window for the simulation settings.

from given source to given destination during that period. Input data in the OD form is useful if we solve dynamic traffic assignment (DTA) problem. Our goal is to make Aurora capable of processing input in OD form. This involves adjustments to the current configuration XML schema (Appendix B).

- So far, "travel time" through a link or path refers to the *instantaneous* travel time as opposed to *actual* travel time. Instantaneous travel time is the travel time that would be experienced if the traffic speed in each link of the path were to stay constant assuming values at current time step. It can be computed every time step as simulation runs. Actual travel time can be only computed after the whole simulation data becomes available. Suppose, we start at a source link 0 at time step k_0 and this link has a queue

$q(k_0)$. The time step at which we arrive at link 1 is

$$k_1 = k_0 + N_0, \tag{5.12}$$

with

$$N_0 = \mathbf{argmax}_k \left\{ \sum_{k'=0}^{k-1} f_d(k_0 + k')\Delta t \leq q(k_0) \right\}, \tag{5.13}$$

where Δt is a sampling period, and $f_d(k_0 + k')$ is the flow leaving link 0 at time step $k_0 + k'$.

If we arrive at link i ($i > 0$) of our path at time step k_i, the actual travel time through this link will be

$$T_i(k_0) = N_i \Delta t, \tag{5.14}$$

with

$$N_i = \mathbf{argmax}_k \left\{ \sum_{k'=0}^{k-1} V_i(k_i + k')\Delta t \leq \Delta x_i \right\}, \tag{5.15}$$

where Δx_i is the length of i-th link, and $V_i(k_i + k')$ is the average traffic speed on i-th link at time step $k_i + k'$.

Given (5.12)-(5.15), arrival time at link $(i+1)$ is $k_{i+1} = k_i + N_i$. Then the total travel time over the path through links $i = 0, 1, \cdots M$ is

$$T(k_0) = \Delta t \sum_{i=0}^{M} N_i, \tag{5.16}$$

with N_i determined from (5.13) and (5.15).

- Last, but not least is how to store the simulation data.

  In a file? But then this file will be too big, and the simulator will require a lot of memory while running. It is an obstacle to making the system truly scalable.

In a database? Makes it easy to export selected data, but raises a question about database maintenance and sharing simulation data.

To tackle this problem some experiments measuring the system performance have to be made, and the TOPl group should decide on the approach to be chosen.

Configuration module.

The configuration module is a cornerstone of the framework, without which Aurora remains handicapped. At this point, it is impossible to make the construction of configuration files completely automatic. The main reason is the large variety of confiuration sources, each with data in its own format: PeMS with its configuration and measurement data; regional planning agencies with their GIS databases; census and demographic data that determines origin-destination travel patterns; and configuration files for other (microsimulation) packages used by different research groups.

Despite the necessary manual intervention in the process of configuration building, the TOPl group decided to automate it where possible. The focus is on GIS databases from regional planning agencies (SANDAG, MTC) as the most consistent and comprehensive sources of information about road geometry. Once the road network is in place, the system must be calibrated. Fundamental diagrams for freeway links can be estimated from PeMS data. For arterials, some best guess default values have to be used. Finally, demand profiles and split ratios, or origin-destination matrices, must be generated using PeMS, census and survey data.

Eventually, PeMS will be collecting data from arterials as well as from freeways, with road configurations coming directly from GIS databases. For California, it will then become a unique source of road geometry together with density-flow data needed for fundamental diagram estimation, and demand profiles and split ratios—everything needed for Aurora configuration files, making the automatic creation of such files possible. Until then, the configuration module is intended for TOPl internal use only, while selected files from our configuration library are provided as part of the Aurora product release.

Analysis module.

Analysis module, currently nonexistent, can be thought of as a collection of special purpose traffic applications relying on simulation data. These could include performance comparison between two or more simulations; shortest path calculation based on actual travel time; demand management; fee computation for tolled lanes or roads, etc.

One of TOPl's first goals is to implement the *dynamic trip assignment* (DTA) application. We have a model of corridor comprising freeways and arterials. The corridor is modeled as a dynamical system (CTM). Underlying the dynamical system is a road network consisting of nodes and links. A subset of node-pairs is identified as a set of origin-destination (OD) pairs. Associated with each OD pair are two entities:

1. demand profile—function of time that gives for each t the flow of vehicles that start at the origin at t and wish to travel to the destination;
2. set of paths or routes through the network that start at the origin and end at the destination that a vehicle is likely to take.

A trip assignment is an assignment of all OD demand profiles to paths. In other words, for each OD pair and time t, the assignment specifies how many of the vehicles will travel over each path associated with the OD pair.

Two types of trip assignments are important.

- *User equilibrium* (UE)—if no individual vehicle can reduce its travel time given that everyone else follows the trip assignment (Wardrop's first principle).
- *System optimal* (SO)—if it minimizes the total travel time summed over all demand profiles (Wardrop's second principle).

In general, UE and SO are different, and

$$T_{UE} \geq T_{SO},$$

where T_{UE} denotes total travel time under UE, and T_{SO} denotes total travel time under SO.

Currently, in the literature we can find description of the standard trip assignment problem [22], in which the demand is stationary (does not depend on time) and there are no dynamics: the delay on a link is simply a function of flow on that link. For such problems, both, UE and SO trip assignments are computed. Trip assignment problems using dynamical system (CTM) as a model are presented in [112, 81], but they focus solely on SO, because computing UE is much more difficult.

We would like to be able to dynamically compute both, UE and SO trip assignments, obtaining "cost of anarchy" as (T_{UE}/T_{SO}).

Chapter 6

Research Plan

The main short term research and development goal is to make Aurora a "product" that can be made available to the users outside the TOPl group. This requires finishing the simulator and implementing configuration module. In the initial version, the simulator is missing signal controllers for arterial junctions, calculation of the actual travel time for user specified paths, and proper simulation data storage facility. The configuration module must be developed as a GIS application using layer databases provided by regional planning agencies as input files, with additional focus on a convenient road network editor that allows bulk provisioning of node and link parameters. Enabling the simulator to handle origin-destination matrices instead of demand and split ratio profiles as alternative input data, and developing dynamic traffic assignment application as a first part of the analysis module, can be considered the next step. This summarizes the short term TODO list for Aurora development.

Following the discussion in Chapter 3, we can pose a problem of finding a control strategy that keeps freeway traffic density at desired levels when demands are not known exactly, but only the bounds on demand values are given. CTM can be treated as piecewise affine system whose *reachable set*[1] can be computed using available techniques. The overview of existing methods and tools for reachability analysis is provided in [68]. Depending on the system dimension, it makes sense to use either polytope library of the Multi Parametric Toolbox (MPT) [69] or Ellipsoidal Toolbox (ET) [68][2] to compute reachable sets and devise control strategies under uncertain demands. Another problem to be investigated, is the extension of Chapter 3 results to a network model [40].

In November, 2006, Caltrans launched an ambitious program to design congestion-reducing operational improvements focusing on demand and incident management, besides ramp metering and traveler information.

One way to manage demand is through tolls. The traditional theory of tolls [21, 25] offers little practical guidance. First, it models congestion by static flow-delay curve that does not reflect queues formed on freeways. Second, the theorem that efficiency requires every link in a freeway to be tolled so as to equate private and social marginal cost is not useful in the practical situation in which only some lanes on a freeway link are tolled. As Kelly observes in [65], "if values of time are heterogeneous or if not all links can be tolled, then the mathematical problem [of finding optimum tolls] becomes non-convex and harder: there may be several radically different candidates for the system optimum, with slight changes

[1]Set of states to which the system can be steered using all admissible inputs from a given initial condition in given number of time steps.

[2]Currently, ET is distributed as part of MPT, which has evolved into a repository of hybrid systems control tools.

in the network specification causing one or another to be preferred."

Queues, missing in the traditional model, were introduced as point queues at bottlenecks and analyzed in [107, 96, 17]. The main theorem is that the dead-weight loss from queuing delay can be eliminated by a time-varying toll. But, as noted in [107], point queues do not model the spillover situation in which "the queue backed up from the bottleneck interferes with the flow of traffic not itself intending to use the bottleneck facility." The spillover situation is studied in [16, 72, 109] for a simple network of two origins, one destination and three links, using kinematic waves to model congestion propagation. The analysis exhibits situations in which ramp-metering and tolls both reduce total travel time. It seems worthwhile to study different toll strategies and their impact on freeway performance within Aurora framework [105].

Incident management comprises four steps as identified by the Federal Highway Administration: *incident detection*, *incident response*, *incident clearance*, *traffic management and traveler information*. Studies have shown that approximately 80% of reported incidents are vehicle disablements, with minor accidents accounting for only 10% of these incidents. The rapid clearance of these incidents, therefore, not only reduces motorist delays, but also reduces the probability of secondary incidents which are often more serious than the primary ones.

Aurora is a suitable framework for developing and testing time saving incident detection algorithms as well as investigating the impact of certain incident response strategies on the overall freeway performance.

Since Aurora is designed as multi-purpose framework, road network is only one of its possible applications. The goal is to try it for a different problem. For example, modeling irrigation networks of open-water channels, that are used throughout the world to support agricultural activity. Traditionally, open-water channel dynamics are described by the so-called Saint-Venant equations [26], nonlinear hyperbolic PDE, which represent a mass and momentum balance along the length of each canal. Computational scheme for such equations is discussed in [15], while important aspects of automating large-scale irrigation networks, including channel modeling and control, are considered in [29].

Once it is established in practice that Aurora is a true multi-purpose framework, i.e., it can support infrastructure networks of different types, it would be interesting to learn how these different type networks can interact and influence each other. Such interaction can be used to model the emergency response and evacuation planning for potential disaster areas as in the Minnesota Evacuation Project [12].

Another application of heterogeneous infrastructure would be simultaneous modeling of road network traffic and public transit. It will help to study the influence of public transit on traffic flow patterns and identify areas where the largest impact can be made.

Appendix A

Configuration File Format for CTMSIM

Configuration file for CTMSIM is a `.mat` MATLAB file that contains the following variables.

- `celldata` (`struct`, $1 \times N$) - Array of freeway cell data structures. N is the number of cells.
- `freeway` (`char`) - String with freeway name or description.
- `TS` (`double`, ≥ 0) - Sampling period. Must be no greater than $\min\left(\frac{\texttt{cell length}}{\texttt{free flow speed}}\right)$.
- `plotTS` (`double`, $\geq$`TS`) - Time period for plotting simulated data.
- `timeout` (`double`, ≥ 0) - Time interval between plot refreshing.
- `initialDensities` (`double`, $N \times 1$, ≥ 0) - Vector of initial densities.

- `inflow` (`double`, ≥ 0) - Flow entering the first cell through mainline.
- `outflow` (`double`, ≥ 0) - Flow that is allowed to leave the last cell through mainline.
- `taxislim` (`double`, ≥ 0) - Upper limit for time axis in time series plots measured in plotted time steps.
- `densityCMF` (`double`, 1024×3) - Color map for time series density contour.
- `flowCMF` (`double`, 1024×3) - Color map for time series flow contour.
- `orflowCMF` (`double`, 1024×3) - Color map for time series on-ramp flow contour.
- `orqueueCMF` (`double`, 1024×3) - Color map for time series on-ramp queue contour.
- `frflowCMF` (`double`, 1024×3) - Color map for time series off-ramp flow contour.
- `frbetaCMF` (`double`, 1024×3) - Color map for time series off-ramp split ratio contour.
- `speedCMF` (`double`, 1024×3) - Color map for time series speed contour.
- `vhtCMF` (`double`, 1024×3) - Color map for time series VHT contour.
- `vmtCMF` (`double`, 1024×3) - Color map for time series VMT contour.
- `yoColorRatio` (`double`, $[0, 1]$) - Array of two numbers that indicate fractions of capacity the flow must achieve in free flow or congested mode to make cell color yellow or orange.
- `demandProfile` (`double`, $K \times (N + 2)$, ≥ 0) - Array of on-ramp demand values for K samples.

- `demandTS` (`double`, $\geq$`TS`) - Sampling time for the on-ramp demands in `demandProfile`.
- `betaProfile` (`double`, $K \times (N+2)$, ≥ 0) - Array of off-ramp split ratios for K samples.
- `betaTS` (`double`, $\geq$`TS`) - Sampling time for the split ratios in `betaProfile`.
- `frflowProfile` (`double`, $K \times (N+2)$, ≥ 0) - Array of off-ramp flows for K samples.
- `frflowTS` (`double`, $\geq$`TS`) - Sampling time for the off-ramp flows in `frflowProfile`.
- `maxSimStep` (`double`, ≥ 0) - Maximum simulation step.
- `maxSimTime` (`double`, ≥ 0) - Maximum simulation time.
- `dataFile` (`char`) - Name of the file where simulation data is to be saved.

The only mandatory variable that must be present in a configuration file is `celldata` - array of freeway cell data structures. Other configuration variables are optional. Variables `maxSimStep`, `maxSimTime` and `dataFile` make difference only when CTMSIM runs in batch mode.

Description of the cell data structure follows.

- `cell.PMstart` (`double`, ≥ 0) - Post mile at cell start.
- `cell.PMend` (`double`, ≥ 0) - Post mile at cell end. Condition `cell.PMstart` > `cell.PMend` implies that traffic moves from right to left.
- `cell.lanes` (`double`, ≥ 0) - Number of lanes. Auxiliary lanes are represented as fractions.

- `cell.FDfmax` (`double`, ≥ 0) - Maximum capacity.
- `cell.FDrhocrit` (`double`, ≥ 0) - Critical density.
- `cell.FDrhojam` (`double`, $\geq$`cell.FDrhojam`) - Jam density. All three fundamental diagram parameters, `cell.FDfmax`, `cell.FDrhocrit` and `cell.FDrhojam` are considered to be *total*, not per lane.
- `cell.ORname` (`char`) - Name of the on-ramp. Empty, if the cell has no on-ramp.
- `cell.ORlanes` (`double`, ≥ 0) - Number of on-ramp lanes.
- `cell.ORflow` (`double`, ≥ 0) - On-ramp flow.
- `cell.ORfmax` (`double`, ≥ 0) - On-ramp capacity.
- `cell.ORqsize` (`double`, ≥ 0) - On-ramp queue size.
- `cell.ORgamma` (`double`, ≥ 0) - On-ramp flow blending coefficient (default: 1).
- `cell.ORxi` (`double`, ≥ 0) - On-ramp flow allocation parameter (default: 1).
- `cell.ORknob` (`double`, ≥ 0) - Coefficient that adjusts on-ramp demand.
- `cell.ORmlcontroller` (`struct`) - On-ramp mainline controller structure (default: `null`).
- `cell.ORqcontroller` (`struct`) - On-ramp queue controller structure (default: `null`).
- `cell.FRname` (`char`) - Name of the off-ramp. Empty, if the cell has no off-ramp.
- `cell.FRlanes` (`double`, ≥ 0) - Number of off-ramp lanes.

- `cell.FRbeta` (`double`, ≥ 0) - Off-ramp split ratio.
- `cell.FRfmax` (`double`, ≥ 0) - Off-ramp capacity.
- `cell.FRknob` (`double`, ≥ 0) - Coefficient that adjusts off-ramp flow.

On-ramp mainline controller structures are expected to have five mandatory fields, which are

- `mlcontroller.id` (`int`) - Controller identifier. Must correspond to the number in the mainline controller list (see Section 4.2.3).
- `mlcontroller.name` (`char`) - String with controller name. Must be nonempty.
- `mlcontroller.TS` (`double`, ≥ 0) - Time period at which controller must be invoked.
- `mlcontroller.Cmin` (`double`, ≥ 0) - Minimum flow recommended by the controller (default: 0).
- `mlcontroller.Cmax` (`double`, ≥ 0) - Maximum flow allowed by the controller.

Other structure fields can be optionally defined by the user.

On-ramp queue controller structures are expected to have two mandatory fields, which are

- `qcontroller.id` (`int`) - Controller identifier. Must correspond to the number in the queue controller list (see Section 4.2.3).
- `qcontroller.name` (`char`) - String with controller name. Must be nonempty.

Other structure fields can be optionally defined by the user.

Appendix B

XML Schema for Aurora Configuration

```
<?xml version="1.0" encoding="UTF-8" ?>

<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">

  <xs:element name="AuroraHWC">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="network" />
        <xs:element ref="settings" />
        <xs:element ref="DemandProfile" />
        <xs:element ref="EventList" />
      </xs:sequence>
    </xs:complexType>
  </xs:element>

  <xs:element name="network">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="description" />
        <xs:element ref="MonitorList" />
        <xs:element ref="NodeList" />
```

```
      <xs:element ref="LinkList" />
      <xs:element ref="ODList" />
    </xs:sequence>
    <xs:attribute name="name" type="xs:string" use="required" />
    <xs:attribute name="controlled" type="xs:boolean" use="required" />
    <xs:attribute name="class" type="xs:string" use="required" />
    <xs:attribute name="top" type="xs:boolean" use="required"
                                              default="false" />
    <xs:attribute name="tp" type="xs:decimal" use="required" />
    <xs:attribute name="id" type="xs:integer" use="required" />
  </xs:complexType>
</xs:element>

<xs:element name="settings">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="display" />
    </xs:sequence>
  </xs:complexType>
</xs:element>

<xs:element name="DemandProfile" minOccurs="0" maxOccurs="1">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="demand" maxOccurs="unbounded" />
    </xs:sequence>
  </xs:complexType>
</xs:element>

<xs:element name="EventList" minOccurs="0" maxOccurs="1">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="event" maxOccurs="unbounded" />
    </xs:sequence>
  </xs:complexType>
</xs:element>

<xs:element name="description" type="xs:string" />

<xs:element name="MonitorList" minOccurs="0" maxOccurs="1">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="monitor" minOccurs="0" maxOccurs="unbounded" />
    </xs:sequence>
```

```
  </xs:complexType>
</xs:element>

<xs:element name="NodeList">
  <xs:complexType>
    <xs:sequence>
      <xs:choice>
        <xs:element ref="node" maxOccurs="unbounded" />
        <xs:element ref="network" maxOccurs="unbounded" />
      </xs:choice>
    </xs:sequence>
  </xs:complexType>
</xs:element>

<xs:element name="LinkList">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="link" maxOccurs="unbounded" />
    </xs:sequence>
  </xs:complexType>
</xs:element>

<xs:element name="ODList" minOccurs="0" maxOccurs="1">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="od" maxOccurs="unbounded" />
    </xs:sequence>
  </xs:complexType>
</xs:element>

<xs:element name="monitor">
  <xs:complexType mixed="true">
    <xs:attribute name="name" type="xs:string" use="required" />
    <xs:attribute name="class" type="xs:string" use="required" />
    <xs:attribute name="id" type="xs:integer" use="required" />
  </xs:complexType>
</xs:element>

<xs:element name="node">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="description" />
      <xs:element ref="outputs" />
      <xs:element ref="inputs" />
```

```
      <xs:element ref="position" />
    </xs:sequence>
    <xs:attribute name="name" type="xs:string" use="required" />
    <xs:attribute name="class" type="xs:string" use="required" />
    <xs:attribute name="id" type="xs:integer" use="required" />
  </xs:complexType>
</xs:element>

<xs:element name="link">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="begin" minOccurs="0" maxOccurs="1" />
      <xs:element ref="end" minOccurs="0" maxOccurs="1" />
      <xs:element ref="fd" />
      <xs:element ref="density" />
      <xs:element ref="dynamics" />
      <xs:element ref="position" />
      <xs:element ref="demand" />
      <xs:element ref="qmax" />
    </xs:sequence>
    <xs:attribute name="lanes" type="xs:decimal" use="required" />
    <xs:attribute name="length" type="xs:decimal" use="required" />
    <xs:attribute name="class" type="xs:string" use="required" />
    <xs:attribute name="id" type="xs:integer" use="required" />
  </xs:complexType>
</xs:element>

<xs:element name="od">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="PathList" />
    </xs:sequence>
    <xs:attribute name="begin" type="xs:integer" use="required" />
    <xs:attribute name="end" type="xs:integer" use="required" />
    <xs:attribute name="class" type="xs:string" use="required" />
  </xs:complexType>
</xs:element>

<xs:element name="outputs">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="output" maxOccurs="unbounded" />
    </xs:sequence>
  </xs:complexType>
```

```
</xs:element>

<xs:element name="inputs">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="input" maxOccurs="unbounded" />
    </xs:sequence>
  </xs:complexType>
</xs:element>

<xs:element name="position">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="point" maxOccurs="unbounded" />
    </xs:sequence>
  </xs:complexType>
</xs:element>

<xs:element name="begin">
  <xs:complexType>
    <xs:attribute name="id" type="xs:integer" use="required" />
  </xs:complexType>
</xs:element>

<xs:element name="end">
  <xs:complexType>
    <xs:attribute name="id" type="xs:integer" use="required" />
  </xs:complexType>
</xs:element>

<xs:element name="fd">
  <xs:complexType>
    <xs:attribute name="densityCritical" type="xs:decimal"
                                         use="required" />
    <xs:attribute name="flowMax" type="xs:decimal" use="required" />
    <xs:attribute name="densityJam" type="xs:decimal" use="required" />
  </xs:complexType>
</xs:element>

<xs:element name="density" type="xs:decimal" />

<xs:element name="dynamics">
  <xs:complexType>
  <xs:attribute name="class" type="xs:string" use="required" />
```

```
  </xs:complexType>
</xs:element>

<xs:element name="demand">
  <xs:complexType mixed="true">
    <xs:attribute name="knob" type="xs:decimal" use="required" />
    <xs:attribute name="tp" type="xs:decimal" use="optional" />
    <xs:attribute name="id" type="xs:integer" use="optional" />
  </xs:complexType>
</xs:element>

<xs:element name="qmax" type="xs:decimal" />

<xs:element name="PathList">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="path" maxOccurs="unbounded" />
    </xs:sequence>
  </xs:complexType>
</xs:element>

<xs:element name="output">
  <xs:complexType>
    <xs:attribute name="id" type="xs:integer" use="required" />
  </xs:complexType>
</xs:element>

<xs:element name="input">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="splitratios" />
      <xs:element ref="controller" />
    </xs:sequence>
    <xs:attribute name="id" type="xs:integer" use="required" />
  </xs:complexType>
</xs:element>

<xs:element name="path">
  <xs:complexType mixed="true">
    <xs:attribute name="name" type="xs:string" use="required" />
    <xs:attribute name="class" type="xs:string" use="required" />
  </xs:complexType>
</xs:element>
```

```
<xs:element name="display">
  <xs:complexType>
    <xs:attribute name="timeMax" type="xs:decimal" use="required" />
    <xs:attribute name="tsMax" type="xs:integer" use="required" />
    <xs:attribute name="timeout" type="xs:integer" use="required" />
    <xs:attribute name="tp" type="xs:decimal" use="required" />
  </xs:complexType>
</xs:element>

<xs:element name="event">
  <xs:complexType>
    <xs:choice>
      <xs:element ref="demand" />
      <xs:element ref="description" />
      <xs:element ref="fd" />
      <xs:element ref="srm" />
    </xs:choice>
    <xs:attribute name="tstamp" type="xs:decimal" use="required" />
    <xs:attribute name="neid" type="xs:integer" use="required" />
    <xs:attribute name="enabled" type="xs:boolean" use="required"
                                                    default="true" />
    <xs:attribute name="class" type="xs:string" use="required" />
  </xs:complexType>
</xs:element>

<xs:element name="point">
  <xs:complexType>
    <xs:attribute name="x" type="xs:decimal" use="required" />
    <xs:attribute name="y" type="xs:decimal" use="required" />
    <xs:attribute name="z" type="xs:decimal" use="required" />
  </xs:complexType>
</xs:element>

<xs:element name="splitratios">
  <xs:complexType mixed="true" />
</xs:element>

<xs:element name="controller">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="parameter" minOccurs="0" maxOccurs="unbounded" />
      <xs:element ref="limits" />
      <xs:element ref="qcontroller" />
    </xs:sequence>
```

```
      <xs:attribute name="class" type="xs:string" use="required" />
      <xs:attribute name="tp" type="xs:decimal" use="required" />
    </xs:complexType>
  </xs:element>

  <xs:element name="parameter">
    <xs:complexType>
      <xs:attribute name="name" type="xs:string" use="required" />
      <xs:attribute name="value" type="xs:string" use="required" />
    </xs:complexType>
  </xs:element>

  <xs:element name="limits">
    <xs:complexType>
      <xs:attribute name="cmin" type="xs:decimal" use="required" />
      <xs:attribute name="cmax" type="xs:decimal" use="required" />
    </xs:complexType>
  </xs:element>

  <xs:element name="qcontroller">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="parameter" minOccurs="0" maxOccurs="unbounded" />
      </xs:sequence>
      <xs:attribute name="class" type="xs:string" use="required" />
    </xs:complexType>
  </xs:element>

  <xs:element name="srm">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="splitratios" />
      </xs:sequence>
    </xs:complexType>
  </xs:element>

</xs:schema>
```

Bibliography

[1] PeMS homepage: http://pems.eecs.berkeley.edu.

[2] TRANSYT homepage: http://www.trlsoftware.co.uk.

[3] FREQ12 homepage: http://www.its.berkeley.edu/research/freq.html.

[4] NETCELL homepage: http://www.its.berkeley.edu/research/netcell.html.

[5] PX-Metacor homepage: http://www.phoenix-isi.fr/en/str/str_main.html.

[6] SATURN homepage: http://www.saturnsoftware.co.uk.

[7] DYNASMART homepage: http://128.8.138.184.

[8] CTMSIM homepage: http://path.berkeley.edu/topl/software.html.

[9] Aurora presentation:
http://path.berkeley.edu/topl/pp/070427_AlexKurzhanskiy__AuroraIntro.pdf.

[10] Java Universal Network/Graph (JUNG) framework: http://jung.sourceforge.net.

[11] JFreeChart homepage: http://www.jfree.org/jfreechart.

[12] Metro Evacuation Project homepage: http://www.spatial.cs.umn.edu/Project/metro-evac/.

[13] C. D. Alecsandru. *A stochastic mesoscopic cell-transmission model for operational analysis of large-scale transportation networks.* PhD thesis, Louisiana State University, 2006.

[14] E. Almasri and B. Friedrich. Online offset optimization in urban networks based on cell transmission model. *ITS Hanover*, 2005.

[15] S. Amin, A. M. Bayen, L. El Ghaoui, and S. Sastry. Robust feasibility for control of water flow in a reservoir-canal system. *Conference on Decision and Control*, 2007. Submitted.

[16] R. Arnott, A. de Palma, and R. Lindsey. Properties of dynamic traffic equilibrium involving bottlenecks. *Transportation Science*, 27(2):148–160, 1993.

[17] R. Arnott, A. de Palma, and R. Lindsey. A structural model of peak-period congestion: a traffic bottleneck with elastic demand. *American Economic Review*, 83(1):161–179, 1993.

[18] A. Aw and M. Rascle. Resurrection of 'second order' models of traffic flow. *SIAM Journal of Applied Mathematics*, 63(1):916–938, 2000.

[19] A. Balogh and M. Krstic. Stability of partial difference equations governing control gains in infinite-dimensional backstepping. *Systems and Control Letters*, 51, 2004.

[20] A. M. Bayen, R. Raffard, and C. J. Tomlin. Adjoint-based constrained control of

eulerian transportation networks: Application to air traffic control. *IEEE Proceedings of the American Control Conference*, 2004.

[21] M. Beckmann, C. B. McGuire, and C. B. Winstein. *Studies in the economics of transportation.* Yale University Press, 1956.

[22] P. Bergendorff, D. W. Hearn, and M. V. Ramana. Congestion toll pricing of traffic networks. *Network Optimization*, 450, 1997.

[23] T. Bewley, R. Temam, and M. Ziane. A general framework for robust control in fluid mechanics. *Physica D*, 138, 2000.

[24] P. Biswas, P. Grieder, and M. Morari. A survey on stability analysis of discrete-time piecewise affine systems. *Proceedings of the 16th IFAC World Congress*, 2005.

[25] D. E. Boyce, H. Mahmassani, and A. Nagurney. A reptrospective on Beckmann, McGuire and Winstein, Studies in the Economics of Transportation. *Papers in Regional Science*, 84(4):85–103, 2005.

[26] W. Brutsaert. The Saint-Venant equations experimentally verified. *Journal of Hydrology*, pages 1387–1401, 1971.

[27] C. Buisson, J.-P. Lebacque, and J. B. Lesort. STRADA, a discretized macroscopic model of vehicular traffic flow in complex networks based on the Godunov scheme. *CESA'96 IMACS Multiconference*, 1996.

[28] C. Buisson, J.-P. Lebacque, J. B. Lesort, and H. Mongeot. The STRADA dynamic assignment model. *Proceedings of the 1996 ITS Conference*, 1996.

[29] M. Cantoni, E. Weyer, Y. Li, S. K. Ooi, I. Mareels, and M. Ryan. Control of large-scale irrigation networks. *Proceedings of the IEEE*, 95(1):75–91, 2007.

[30] R. Cayford, W.-H. Lin, and C. F. Daganzo. The NETCELL simulation package: technical description. Technical Report UCB-ITS-PRR-97-23, California PATH, University of California, Berkeley, CA 94720, 1997.

[31] C. Chen, J. Kwon, A. Skabardonis, and P. Varaiya. Detecting errors and imputing missing data for single loop surveillance systems. *Transportation Research Record*, (1855):160–167, 2003.

[32] M.-L. Chen and D. Georges. Nonlinear optimal control of an open-channel hydrolic system based on an infinite-dimensional model. *Proceedings of the 38th IEEE Conference on Decision and Control*, 1999.

[33] L. Chu and H. X. Liu. PARAMICS plug-in document: SWARM ramp metering control. Technical report, California PATH, University of California, Berkeley, CA 94720, 2003.

[34] S. S. Collis, K. Ghayour, M. Heikenschloss, M. Ulbrich, and S. Ulbrich. Optimal control of unsteady compressible viscous flows. *International Journal for Numercal Methods in Fluids*, 40, 2002.

[35] R. M. Colombo. Hyperbolic phase transitions in traffic flow. *SIAM Journal of Applied Mathematics*, 63(2):708–721, 2002.

[36] R. M. Colombo. On a 2x2 hyperbolic traffic model. *Math. Comput. Modelling*, 35:683–688, 2002.

[37] J.-M. Coron, B. d'Andréa Novel, and G. Bastin. A strict Lyapunov function for boundary control of hyperbolic systems of conservation laws. *Proceedings of the 43rd IEEE Conference on Decision and Control*, 2004.

[38] F. Cuzzola and M. Morari. An LMI approach for H_∞ analysis and control of discrete-time piecewise affine systems. *International Journal of Control*, 75(16-17), 2002.

[39] C. F. Daganzo. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research, B*, 28(4):269–287, 1994.

[40] C. F. Daganzo. The cell transmission model II: Network traffic. *Transportation Research, B*, 29(2):79–93, 1995.

[41] C. F. Daganzo. Requiem for second-order fluid approximations of traffic flow. *Transportation Research, B*, 29:277–286, 1995.

[42] C. F. Daganzo and W.-H. Lin. The spatial evolution of queues during the morning commute in a single corridor. Technical Report UCB-ITS-PWP-93-7, California PATH, University of California, Berkeley, CA 94720, 1993.

[43] J. de Halleux, C. Prieur, J.-M. Coron, B. d'Andréa Novel, and G. Bastin. Boundary feedback control in networks of open channels. *Automatica*, 2003.

[44] N. Elloumi, H. Haj Salem, and M. Papageorgiou. METACOR: A dynamic macroscopic modeling tool for urban corridor. *TRIennal Symposium on Transport Analysis, Capri*, 1, 1994.

[45] O. Feldman and M. Maher. Optmization of traffic signals using a cell transmission model. *34th Annual Universities' Transport Study Group Conference, Napier University, Edinburgh*, 2002.

[46] M. Garavello and B. Picolli. Traffic flow on networks. *'Applied Mathematics Series', American Institute of Mathematical Sciences*, (1), 2006.

[47] O. Ghattas and J. Bark. Optimal control of two- and three-dimensional incompressible Navier-Stokes flows. *Journal of Computational Physics*, 136, 1997.

[48] M. Giles and N. Pierce. Analytic adjoint solutions for the quasi-one-dimensional Euler equations. *Journal of Fluid Mechanics*, 426, 2001.

[49] E. Godlewski and P.-A. Raviart. *Numerical Approximation of Hyperbolic Systems of Conservation Laws.* Springer-Verlag, 1996.

[50] S. K. Godunov. A difference method for numerical calculation of discontinuous solutions of hydrodynamical equations. *Matematicheskii Sbornik*, 47:271–290, 1959.

[51] G. Gomes and R. Horowitz. Optimal freeway ramp metering using the asymmetric cell transmission model. *Transportation Research, C*, 14(4):244–262, 2006.

[52] G. Gomes, R. Horowitz, A. A. Kurzhanskiy, J. Kwon, and P. Varaiya. Behavior of the Cell Transmission Model and Effectiveness of Ramp Metering. *Transportation Research, C*, 2007. Submitted.

[53] B. Greenshields. A study of highway capacity. *Proceedings of the 14th annual meeting of the Highway Research Board*, 1934.

[54] D. Helbing. Improved fluid dynamic model for vehicular traffic. *Physical Review E*, pages 3164–3169, 1995.

[55] D. Helbing. Traffic and related self-driven many-particle systems. *Reviews of Modern Physics*, 73:1067–1141, 2001.

[56] M. Hinze and K. Kunisch. Second order methods for optimal control of time-dependent fluid flow. *SIAM Journal on Control and Optimization*.

[57] M. W. Hirsch and H. Smith. Monotone maps: a review. *Journal of Difference Equations and Applications*, 11(4-5):379–398, 2005.

[58] ITT Systems and Sciences. CORSIM user's manual. *Office of Safety and Traffic Operations R&D, Federal Highway Administrtion, Washington, D.C.*, 1998.

[59] D. Jacquet. *Macroscopic Traffic Modeling and Control of Conservation Laws*. PhD thesis, Institut National Plytechnique de Grenoble, 2006.

[60] A. Jameson. Optimum aerodynamic design using cfd and control theory. *12th AIAA Computational Fluid Dynamics Conference*, 1995.

[61] A. Jameson. Aerodynamic shape optimization using the adjoint method. *Lecture Series at the Von Karman Institute, Brussels, Belgium*, 2003.

[62] A. Jameson, L. Martinelli, and N. Pierce. Optimum aerodynamic design using Navier-Stokes equations. *Theoretical Computational Fluid Dynamics*, 10, 1998.

[63] R. Jayakrishnan, H. S. Mahmassani, and T.-Y. Hu. An evaluation tool for advanced

traffic information and management systems in urban networks. *Transportation Research, C*, 2(3):129–147, 1994.

[64] M. Johansson and A. Rantzer. Computation of piecewise quadratic Lyapunov functions for hybrid systems. *IEEE Transactions on Automatic Control*, 43(4), 1998.

[65] F. P. Kelly. Road pricing. *Ingenia*, 29:36–40, 2006.

[66] A. Kotsialos, M. Papageorgiou, and A. Messmer. Optimal co-ordinated and integrated motorway network traffic control. *Proceedings of the 14th International Symposium of Transportation and Traffic Theory, Jerusalem*, pages 621–644, 1999.

[67] M. Krstic. On global stabilization of Burgers equation by boundary control. *Systems and Control Letters*, 37, 1999.

[68] A. A. Kurzhanskiy and P. Varaiya. Ellipsoidal Toolbox. Technical Report UCB/EECS-2006-46, EECS Department, University of California, Berkeley, 2006. Online: http://www.eecs.berkeley.edu/~akurzhan/ellipsoids.

[69] M. Kvasnica, P. Grieder, M. Baotić, and M. Morari. Multi Parametric Toolbox (MPT). *Lecture Notes in Computer Science*, 2993:448–462, 2004.

[70] J. Kwon. Joint estimation of the traffic speed and mean vehicle length from single-loop detector data. *82nd Annual Meeting of the Transportation Research Board, Washington, D.C., USA*, 2003.

[71] J. Kwon and P. Varaiya. California's bottlenecks. 2005. Submitted to the GoCalifornia Expert Review Panel.

[72] A. Lago and C. F. Daganzo. Spillovers, merging traffic and the morning commute. *Transportation Research, B*, 41:670–683, 2007.

[73] J.-P. Lebacque. Simulation semi-macroscopique des réseaux urbains. Technical report, INRETS, Arqueil, France, 1983.

[74] J.-P. Lebacque. The Godunov scheme and what it means for first order traffic flow models. *Proceedings of the 13th International Symposium of Transportation and Traffic Theory, Lyon*, pages 647–677, 1996.

[75] R. J. LeVeque. *Numerical Methods for Conservation Laws.* Birkhäuser, 1992.

[76] H. Lieu and A. J. Santiago. CORFLO: An integrated traffic simulation system for corridors. *Traffic Management, Proceedings of the Engineering Foundation Conference, Palm Coast, Florida*, 1991.

[77] M. Lighthill and G. Witham. On kinematic waves I: Flow movement in long rivers. II: A theory of traffic flow on long crowded roads. *Proc. Royal Society of London, Part A*, 229(1178):281–345, 1955.

[78] W.-H. Lin and D. Ahanotu. Validating the basic cell transmission model on a single freeway link.

[79] J. L. Lions. *Optimal control of systems governed by partial differential equations.* Springer-Verlag, 1971.

[80] G. Liu, A. S. Lyrintzis, and P. G. Michalopoulos. Improved high-order model for freeway traffic flow. *Transportation Research Record*, (1644):37–46, 1998.

[81] H. K. Lo. A cell-based traffic control formulation: Strategies and benefits of dynamic timing plans. *Transportation Science*, 35(2):148–164, 2001.

[82] H. K. Lo, E. Chang, and Y. C. Chan. Dynamic network traffic control. *Transportation Research, A*, 35:721–744, 2001.

[83] H. K. Lo and A. H. F. Chow. Control strategies for oversaturated traffic. *Journal of Transportation Engineering*, 130(4):721–744, 2004.

[84] A. S. Lyrintzis, G. Liu, and P. G. Michalopoulos. Development and comparative evaluation of high-order traffic flow models. *Transportation Research Record*, (1547):174–183, 1994.

[85] A. Messmer and M. Papageorgiou. METANET: A macroscopic simulation program for motorway networks. *Traffic Engineering and Control*, 31(8/9):466–470, 1990.

[86] L. Munõz, X. Sun, R. Horowitz, and L. Alvarez. Traffic density estimation with the cell transmission model. *Proceedings of the 2003 American Control Conference*, pages 3750–3755, 2003.

[87] L. Munõz, X. Sun, D. Sun, G. Gomes, and R. Horowitz. Methodological calibration of the cell transmission model. *Proceedings of the 2004 American Control Conference*, pages 798–803, 2004.

[88] Y. Nie. *A variational inequality approach for inferring dynamic origin-destination travel demands.* PhD thesis, University of California, Davis, 2006.

[89] M. Papageorgiou. Dynamic modelling, assignment and route guidance in traffic networks. *Transportation Research, B*, 23:29–48, 1989.

[90] M. Papageorgiou, H. Hadj-Salem, and H. M. Blosseville. ALINEA: A local feedback control law for onramp metering. *Transportation Research Record*, (1320), 1991.

[91] H. J. Payne. Models of freeway traffic and control. *Simulation Council Proceedings*, 1, 1971.

[92] H. J. Payne. FREFLO: A macroscopic simulation model of freeway traffic. *Transportation Research Record*, (722):68–77, 1979.

[93] P. Richards. Shock waves on the highway. *Operations Research*, 4(1):42–51, 1956.

[94] B. Sanders and N. Katopodes. Sensitivity analysis of shallow water flow by adjoint method. *Journal of Engineering Mechanics*, 126(9):909–919, 2000.

[95] A. Skabardonis and N. Geroliminis. Real-time estimation of travel times on signalized arterials. *Proceedings of 16th International Symposium on Transportation and Traffic Theory, University of Marland, Elsevier*, pages 387–406, 2005.

[96] K. A. Small. The schedulig of consumer activities: work trips. *American Economic Review*, 72(3):467–479, 1982.

[97] S. Smulders. Control of freeway traffic flow by variable speed signs. *Transportation Research, B*, 24(2):111–132, 1990.

[98] X. Sun and R. Horowitz. A set of new traffic-responsive ramp-metering algorithms and microscopic simulation results. *Transportation Research Record*, 2006.

[99] System Metrics Group, Inc. Transportation Management System (TMS) Master Plan. 2004.

[100] J. Tan. Calibration of cell transmission model to morning commute traffic on I210-W. Master's thesis, University of California, Berkeley, 2003.

[101] Transportation Research Bord. Highway capacity manual. *National Research Council, Washington, D.C.*, 2000.

[102] M. F. A. M. Van Maarseveen. *The Theory of Martingales n Stochastic Systems Theory – Surveillance and Control of Freeway Traffic Flow.* PhD thesis, University of Twente, Enschede, 1982.

[103] D. Van Vliet. SATURN - a modern traffic assignment model. *Traffic Engineering and Control*, 23(4), 1982.

[104] D. Van Vliet. Fundamental requirements for full scale route guidance - recent developments to SATURN. *Memorandum, University of Leeds*, 1990.

[105] P. Varaiya. Freeway congestion, ramp metering, and tolls. Presented at Royal Discussion Society Discussion on 'Networks: Modeling and Control', September 24-25, 2007.

[106] P. Varaiya. How to measure transportation system performance. 1997. Online: http://pems.eecs.berkeley.edu/Resources.

[107] W. S. Vickery. Congestion theory and transport investment. *American Economic Review*, 59(2):251–260, 1969.

[108] G. Whitham. Linear and nonlinear waves. *Pure and Applied Math., Wiley-Interscience, New York*, 1974.

[109] I. Yperman, S. Logghe, and B. Immers. Dynamic congestion pricing in a network with queue splillover. *12th World Congress on Intelligent Transport Systems, ITS America*, 2005.

[110] H. M. Zhang. A non-equilibrium traffic model devoid of gas-like behavior. *Transportation Research, B*, 36, 2002.

[111] M. Zhang, S. G. Ritchie, and W. W. Recker. Some general results on the optimal ramp control problem. *Transportation Research, C*, 4(2):51–69, 1996.

[112] A. K. Ziliaskopoulos. A linear programming model for the single destination system optimum dynamic traffic assignment problem. *Transportation Science*, 34(1):37–49, 2000.

[113] A. K. Ziliaskopoulos and S. Lee. A cell transmission based assignment-simulation model for integrated freeway/surface street systems. *Transportation Research Record*, (1701):12–23, 1997.