

TRAFFIC FLOW THEORY

TRAFFIC FLOW THEORY

Characteristics, Experimental Methods, and Numerical Techniques

DAIHENG NI

Department of Civil and Environmental Engineering,
University of Massachusetts Amherst
MA, USA



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Butterworth-Heinemann is an imprint of Elsevier



Butterworth Heinemann is an imprint of Elsevier
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
225 Wyman Street, Waltham, MA 02451, USA

Copyright © 2016 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

For information on all Butterworth Heinemann publications
visit our website at <http://store.elsevier.com/>

ISBN: 978-0-12-804134-5



Working together
to grow libraries in
developing countries

www.elsevier.com • www.bookaid.org

הלווייה

PREFACE

For years, I have been thinking about writing an introductory book on traffic flow theory. The main purpose is to help readers who are new to this subject and who do not have much knowledge of mathematics and traffic flow. To serve this purpose, I have tried to make the contents self-contained and assume minimal knowledge of mathematics and traffic flow.

This book is derived from my lecture notes for CEE520 Traffic Flow Theory and Simulation I (formerly offered as CEE590T Traffic Flow Theory on an experimental basis before it was assigned a permanent course number) at the University of Massachusetts Amherst. Hence, the chapters are more like lectures, with focused topics, each of which fits in a class meeting. The book takes a unified perspective on traffic flow modeling and consists of five parts which are coherently connected. Each part is briefly described as follows.

Part I focuses on traffic flow characteristics. It starts with intelligent transportation systems and traffic sensing technologies to illustrate how to quantify traffic flow and collect such data. This is followed by three chapters with in-depth discussion of traffic flow characteristics, on the basis of which their relationships are developed and a few equilibrium traffic flow models are introduced.

Part II is about traffic flow modeling at the *macroscopic* level. The goal is to solve for temporal-spatial evolution of traffic flow characteristics given initial and boundary conditions. The first few chapters provide a jump start on mathematical modeling, especially partial differential equations. With such knowledge, the domain knowledge of traffic flow is integrated into mathematical modeling, resulting in a first-order quasi-linear partial differential equation problem known as the Lighthill, Whitham, and Richards (LWR) model in the traffic flow community. Solutions to the problem are introduced, including a graphical technique that uses the method of characteristics and numerical techniques that involves a few discretization schemes.

Part III is devoted to traffic flow modeling at the *microscopic* level. The emphasis is on drivers' car-following behavior involving operational control in the longitudinal direction. A series of car-following models with differing modeling philosophies and complexity are introduced. To provide

an opportunity to cross-compare the relative performance of these models, a common ground is set up so that these models can demonstrate themselves. Such a process is called benchmarking, and the common ground consists of two scenarios, one microscopic and the other macroscopic. The microscopic scenario is a hypothetical driving process aimed at testing these models under various driving regimes (such as free flow and car following); the macroscopic scenario is a set of empirical data focusing on examining the macroscopic properties of these models (e.g., how their implied fundamental diagrams compare with the observed diagrams).

Part IV extends traffic flow modeling to the *piconoscopic* level. A modeling framework called a driver-vehicle-environment closed-loop system is introduced to capture the ultrafine level of detail of traffic flow. Such a framework involves a driver model, a vehicle model, and the driving environment. The driver model collects and processes information from its vehicle and the driving environment and makes control decisions on motion in longitudinal and lateral directions. The vehicle model executes its driver's control decision and moves dynamically on the road. The driver-vehicle unit constitutes one of the entities in the environment whose dynamic change affects driver control in the next step. As an example of this modeling framework, a simple engine model and further a dynamic interactive vehicle model are proposed, and a field theory is formulated to model the driver.

All things come together in Part V. With the field theory as the basis, a unified perspective can be cast on traffic flow theory. The macroscopic models and microscopic models introduced thus far can be related to each other, all linked directly or indirectly to the field theory. Hence, a unified diagram is constructed to highlight such relations. In addition, benchmarking is done to cross-compare the performance of some of the macroscopic models and microscopic models in the diagram. Further, a multiscale modeling approach is presented which involves traffic flow modeling at four levels of detail—namely, macroscopic, mesoscopic, microscopic, and piconoscopic. The emphasis of multiscale modeling is to ensure modeling consistency—that is, how less detailed models are derived from more detailed models and, conversely, how more detailed models are aggregated to less detailed models. The proposed approach may establish the theoretical foundation for traffic modeling and simulation at multiple scales seamlessly within a single system.

This book is ideal for use by entry-level graduate students in transportation engineering as a textbook for a traffic flow theory course. In addition,

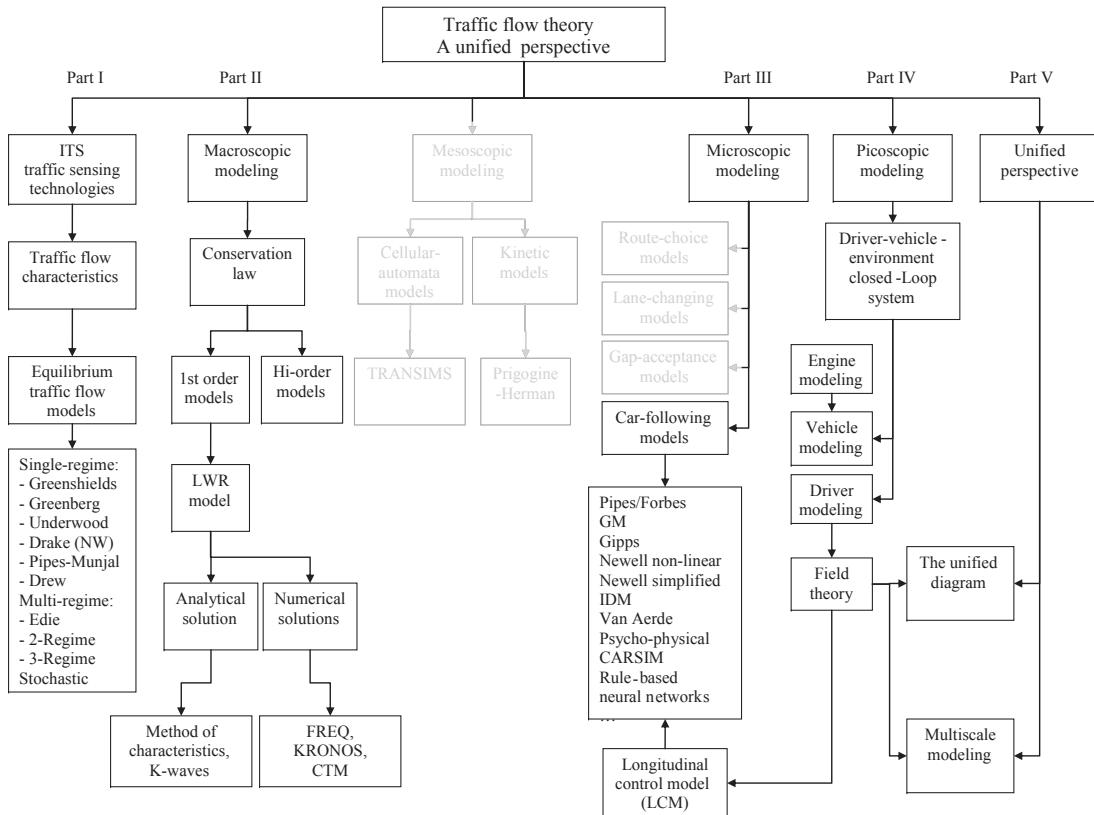
civil engineering juniors and seniors may find some in-depth information about traffic flow fundamentals in this book. Further, applied mathematics majors may find concrete examples of mathematical modeling with specific domain knowledge. Advanced readers are referred to other traffic flow theory books for in-depth coverage; a few of them are as follows:

- G.F. Newell, Theory of Highway Traffic Flow, 1945–1965, Course Notes UCB-ITS-CN-95-1, 1996.
- A.D. May, Traffic Flow Fundamentals, Prentice-Hall, New York, 1989.
- C.F. Daganzo, Fundamentals of Transportation and Traffic Operations, Pergamon-Elsevier, Oxford, UK, 1997.
- N. Gartner, C.J. Messer, A.K. Rathi, Revised Monograph on Traffic Flow Theory: A State-of-the-Art Report, TRB, 2001.
- D.L. Gerlough, M.J. Huber, Traffic Flow Theory—A Monograph, TRB Special Report 165, 1975.
- D.L. Gerlough, D.G. Capelle, An Introduction to Traffic Flow Theory, HRB Special Report 79, 1964.
- D.R. Drew, Traffic Flow Theory and Control, McGraw-Hill, New York, 1968.
- W. Leutzbach, Introduction to the Theory of Traffic Flow, Springer-Verlag, New York, 1988.
- M. Treiber, A. Kesting, Traffic Flow Dynamics, Springer, New York, 2013.
- L. Elefteriadou, An Introduction to Traffic Flow Theory, Springer, New York, 2014.
- B.S. Kerner, Introduction to Modern Traffic Flow Theory and Control, Springer, New York, 2009.

I thank Professor John D. Leonard at Georgia Institute of Technology and Professor Billy M. Williams at North Carolina State University, who introduced me to this field and sparked my interest in traffic flow theory. Thanks also go to former students in my traffic flow theory classes—their insightful discussion and kind encouragement made this work possible.

Finally, I acknowledge my limitations. Though I have tried hard to ensure the quality and accuracy of information, I can make mistakes. Therefore, readers should use this book with discretion.

Daiheng Ni
Amherst, MA
September, 2015



Note: Gray areas are part of traffic flow theory but not covered in this book.

CHAPTER 1

Traffic Sensing Technologies

Safe and efficient operations of transportation systems rely heavily on applications of advanced technologies. As a result, recent decades have witnessed wide applications of communication, sensing, and computing technologies in traffic surveillance, incident detection, emergency response, fleet management, and travel assistance. [Figure 1.1](#) illustrates an example of these technologies at an intersection.

“Intelligent transportation systems” (ITS) refers to efforts that apply information, communication, and sensor technologies to vehicles and transportation infrastructure in order to provide real-time information for road users and transportation system operators to make better decisions. ITS aim to improve traffic safety, relieve traffic congestion, reduce air pollution, increase energy efficiency, and improve homeland security. ITS encompass a suite of measures that address the above objectives: advanced traffic management systems, advanced traveler information systems, advanced public transportation systems, the intelligent vehicle initiative, the commercial vehicle operations program, etc. The recent development of ITS emphasizes the application of dedicated short-range communications in vehicle-to-vehicle and vehicle-to-roadside wireless communications—that is, connected vehicle technology according to the US Department of Transportation.

1.1 TRAFFIC SENSORS

This section describes a few types of traffic sensors that are often employed in ITS and other traffic surveillance and data collection systems. The discussion of each type of sensor focuses on how it works, what traffic data it is capable of collecting, its advantages, and its disadvantages.

1.1.1 Inductive-Loop Detector

Inductive-loop detectors are widely used at intersections with traffic-actuated signals, freeway entrances with automatic ramp metering, highway segments monitored by traffic counting programs, and entrances of gated parking facilities.

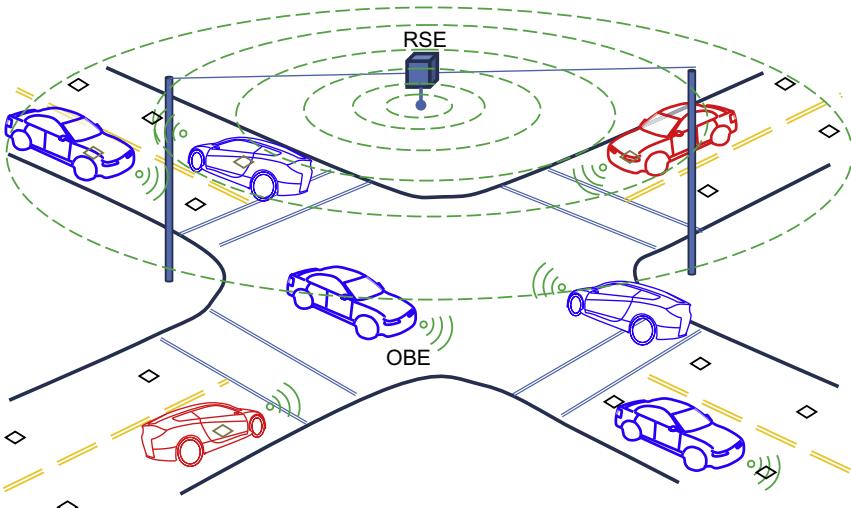


Figure 1.1 An example application of connected vehicles at an intersection.

How It Works

As illustrated in [Figure 1.2](#), an inductive-loop detection system consists of an inductive loop, which is simply a coil of wire embedded in the road's pavement, and a detector, which typically sits in a signal cabinet and links the signal controller to the inductive loop. The detector drives an alternating flow of current through the loop at or below the resonant frequency. All wire conductors carrying an electrical current produce a magnetic field, and the magnetic flux induces the electrical property called inductance. Note that the metal body and frame provide a conductive path for the magnetic field. Therefore, when a vehicle enters the detection zone or crosses the loop, this produces a loading effect, which in turn causes the loop inductance to decrease. The decreased inductance causes the resonant frequency to increase from its nominal value. If the frequency change exceeds the threshold set by the sensitivity setting, the detector module will output a detect signal—that is, an “on” state. Otherwise, the detector does not output a signal—that is, an “off” state.

The output of the detector can be used for many applications. For example, an actuated signal controller relies on the detector output to decide whether a green indication is granted to the approach that is monitored by the detector. As another example, when a vehicle exits a gated parking garage, an inductive loop is able to detect the vehicle in advance so that the

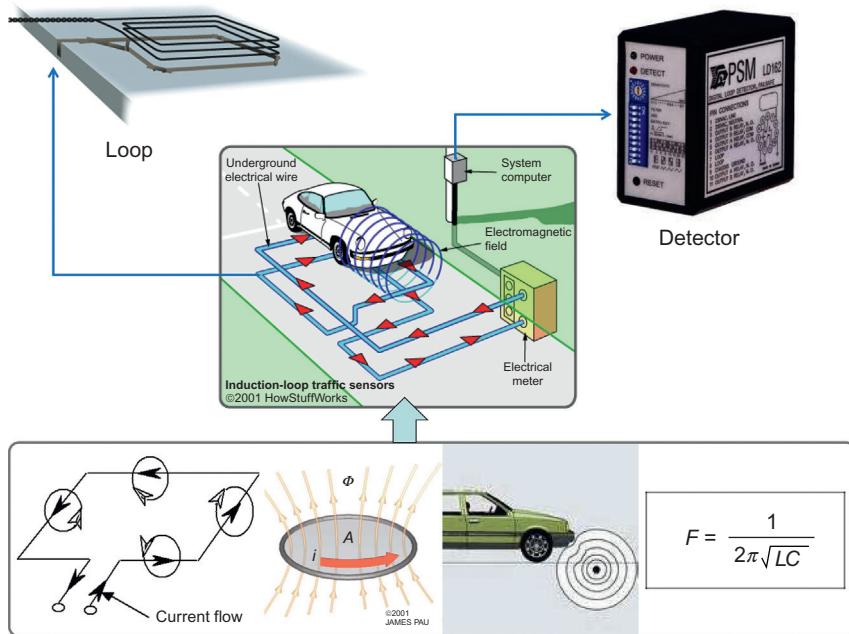


Figure 1.2 An inductive-loop detection system.

gate automatically opens for the vehicle. Yet another innovative application is a red-light-running camera. An intersection with such a system has the detector connected to the signal controller and an overhead camera. As a result, when a vehicle is running a red light, the camera will be triggered and a picture of the vehicle will be taken as evidence of red light violation.

Data Collected

An inductive-loop detector monitors a point of the roadway and is able to collect time-stamped traffic counts with vehicle classification, vehicle instantaneous speed, headway (temporal separation between two consecutive vehicles), on time (time during which the detector outputs an “on” state), etc.

Advantages

An inductive-loop detector is able to monitor traffic on a regular basis (i.e., day-round and year-round) under all weather and lighting conditions.

Disadvantages

Installation of inductive-loop detectors is intrusive to traffic (i.e., the traffic must be interrupted in order to put the loop in the pavement). In addition, setup and maintenance costs of inductive-loop detectors are high. Inductive-loop detectors can fail under some weather conditions, especially snow and ice.

1.1.2 Video Image Processing System

A video image processing system (VIPS) is widely used for traffic surveillance and hence is an essential component of ITS.

How It Works

A VIPS comprises (1) an image capturing system (e.g., a video camera mounted above the roadway that captures real-time images/video streams of the traffic under surveillance), (2) a telecommunication system (e.g., a modem and a telephone line that transmit images/video streams to the image processing system), and (3) an image processing system (e.g., a computer that processes frames of a video clip to extract traffic data).

The left panel in Figure 1.3 illustrates a video camera which is monitoring traffic. The right panel shows an image of roadway traffic (not necessarily a match of the view of the video camera in the left panel) with detection zones set up on the screen. When a vehicle enters a detection zone, the VIPS outputs an “on” signal, which remains until the vehicle exits the detection zone, at which time the VIPS switches to an “off” signal. Multiple detection zones can be set up—for example, one for each lane. Hence, these detection zones constitute a detection station.



Figure 1.3 Video image processing system. (*Photos from <http://www.imagesensing.com/>*)

Data Collected

Similarly to inductive-loop detectors, the VIPS monitors a point of the roadway and is able to collect time-stamped traffic counts with vehicle classification, vehicle instantaneous speed, headway, on time, etc.

Advantages

The VIPS is an automatic system and is able to collect traffic data on a regular basis. Its overhead installation makes this technology nonintrusive to traffic flow. It is flexible in the setting up of detection zones and aggregation intervals. It provides video footage in addition to traffic monitoring.

Disadvantages

The VIPS is expensive and its setup cost is high. It is vulnerable to visual obstruction—for example, inclement weather, shadows, poor-lighting conditions, and strong winds.

1.1.3 Pneumatic Tubes

Pneumatic tubes are portable traffic data collection devices and are ideal for short-term traffic engineering studies.

How It Works

A rubber tube with a diameter of about 1 cm is placed on the surface of a road. When a vehicle passes, the wheel presses the tube and the air inside the tube is pushed away. One end of the tube is connected to a box that contains a membrane and an electrical switch. The air pressure moves the membrane and engages the switch. The other end of the tube has a small opening, to prevent reflection of the air wave. The box counts axles that travel over the tubes and stores the data for later analysis.

Figure 1.4 illustrates how pneumatic tubes are installed: from left to right, a technician is nailing tubes on the road; the technician is programming the data recorder with a laptop computer to collect the desired information; the technician is connecting the pneumatic tubes to the data collector; the installation is complete and the system is collecting traffic data.



Figure 1.4 Installation of pneumatic tubes. (Photos from <http://www.arlingtonva.us/>.)

Data Collected

Rather than collecting traffic counts as in the previous two types of sensors, pneumatic tubes are able to collect time-stamped axle counts, from which vehicle classification, direction of flow, traffic counts, flow, vehicle instantaneous speed, headway, and on time can be inferred.

Advantages

Pneumatic tubes are portable devices for automatic traffic data collection. The cost is moderate, and the system can be reused at other locations. Installation can be done by one or two persons.

Disadvantages

The system has a limited lane coverage and is not intended for use on a regular basis (year-round). The system can be damaged by vehicles or roadway maintenance, causing inaccurate data collection. The system may be intrusive to traffic and nearby properties.

1.1.4 Global Positioning System Receiver

The global positioning system (GPS) is widely used in automotive navigation and traffic engineering studies such as traffic time studies. Many cell phones are equipped with positioning functions, and hence they are considered in the same category as the GPS.

How It Works

The GPS is a satellite-based navigation system made up of a network of 24 satellites placed in orbit by the US Department of Defense. GPS satellites circle Earth twice a day in a very precise orbit and transmit signal information to Earth. GPS receivers take this information and use triangulation to calculate the user's exact location (see [Figure 1.5](#) for an illustration). Essentially, the GPS receiver compares the time when a signal was transmitted by a satellite with the time when it was received. The



Figure 1.5 The global positioning system. (Photos from https://en.wikipedia.org/wiki/Global_Positioning_System.)

time difference tells the GPS receiver how far away the satellite is. Now, with distance measurements from a few more satellites, the receiver can determine the user's position and display it on the unit's electronic map.

If a vehicle carries a GPS receiver on board and it is set up to log GPS signals, it is possible to record the positions of the vehicle and the time when a location is passed as the vehicle moves along the road. Therefore, the vehicle would leave a trace of spatial-temporal points in the time-space diagram, and a curve that connects these points depicts the vehicle's spatial-temporal trajectory. From this trajectory, the motion of this vehicle can be understood.

Data Collected

Vehicle-specific motion data such as instantaneous speed, average running speed, distance traveled, and travel time are collected.

Advantages

GPS has become an affordable technology since one only needs a GPS receiver to receive positioning signals. GPS receivers are simple to install and operate. They work under all weather and lighting conditions.

Disadvantages

GPS receivers provide only vehicle-specific data. Traffic information has to be obtained from all vehicles in the traffic stream. In addition, GPS signals can be obstructed by tall buildings and trees.

1.1.5 Acoustic/Ultrasonic Sensor

Acoustic/ultrasonic sensors can be used for vehicle detection, automotive radar, and assisting vehicle parking.

How It Works

The sensor shoot a beam of sound, like radar, which travels until it hits an object. The sound wave then bounces back and returns to the sensor. The sensor then measures the time it takes the sound wave to travel. Knowing the speed of sound, the sensor outputs the distance between the sensor and the object. In traffic applications, these sensors can be used to count pedestrians and vehicles by knowing the distance between a pedestrian/vehicle and the sensor. In mechanical applications, these sensors can be used to measure fluid levels. The photo in [Figure 1.6](#) shows them installed in the rear of a vehicle as a parking sensor. The sensors measure the distance between the vehicle and an object behind the vehicle, and then



Figure 1.6 Acoustic/ultrasonic sensors. (Photo from <http://autoteksheffield.co.uk/security/parking-sensors/>.)

display a color corresponding to the distance on the dashboard panel. When the display turns red, the driver can stop and is perfectly parked.

Data Collected

The sensor collects the time of sound wave travel, and then converts it to distance.

Advantages

The sensor is inexpensive in general and involves relatively simple hardware.

Disadvantages

The sensor covers only a short range and has slow response times. Accuracy is limited by the surface of the objects. Sound waves may bounce off various surfaces differently, which may throw off readings on the sensor.

1.1.6 Aerial/Satellite Imaging

How It Works

This technology usually requires the use of either manned or unmanned helicopters in the sky to monitor and observe traffic on the ground for data collection purposes. Illustrated in **Figure 1.7**, the helicopter can be used to capture images of the ground, and the images are stored or transmitted to a workstation for analysis. The information obtained includes vehicle counts, vehicle speeds, and traffic density.

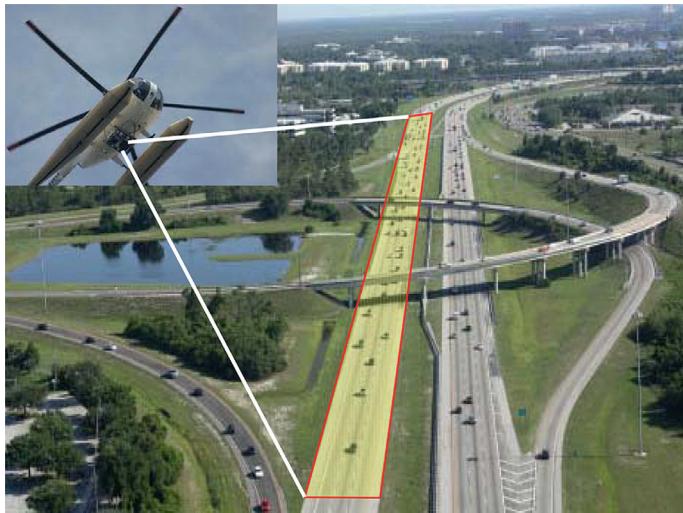


Figure 1.7 Unmanned helicopter as a traffic sensor.

Data Collected

The captured aerial photos contain snapshots of traffic on roadways, from which spatial traffic data such as spacing (i.e., spatial separation between two consecutive vehicles), vehicle counts over a segment of roadway, and traffic density can be obtained. In addition, analysis of consecutive aerial photos may yield information about vehicle speeds and mean traffic speed.

Advantages

Traffic surveillance can be done at high accuracy. There is no need for hardware installation on or near roadways—that is, it is a nonintrusive and noninterruptive technology. It can provide a bird's eye view of system-wide traffic conditions.

Disadvantages

Helicopters are expensive and may require pilots to operate them. It is time-consuming and resource-consuming to collect traffic data. Analysis of aerial photos is complicated—for example aligning aerial photos captured from different angles and extracting traffic data from these photos.

1.1.7 Radio-Frequency Identification Technology

Radio-frequency identification (RFID) is the core technology of many traffic sensors known as transponders (e.g., E-ZPass tags), and is used for automatic vehicle identification, etc.

How It Works

RFID is a technology that uses radio waves to exchange data between a reader and an electronic tag attached to an object for the purpose of identification and tracking. Figure 1.8 illustrates an electronic toll collection system which consists of (1) a transponder on the vehicle, (2) a tag reader antenna at each plaza toll lane, (3) lane controllers that control the lane equipment and track vehicles passing through, and (4) a host computer system. All of the toll plaza controllers are connected to a central database. When a vehicle comes to the toll booth, the tag reader detects the transponder and records its unique ID, the time instant, and other account-related information such as balance and toll paid.

Data Collected

RFID technology is able to record the IDs of equipped vehicles and timestamp the arrival of these vehicles.

Advantages

RFID technology is inexpensive. It does not interrupt traffic.

Disadvantages

RFID only detects equipped vehicles at a point of roadway.

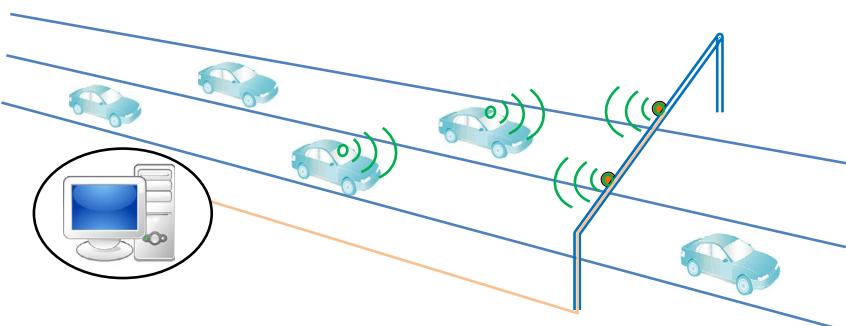


Figure 1.8 Electronic toll collection system.

1.2 TRAFFIC SENSOR CLASSIFICATION

Traffic sensors can be classified in many ways. For example, according to its working principle, a traffic sensor can be a

- *mobile sensor* if it resides in a vehicle and collects data only specific to this vehicle. GPS receivers, acoustic/ultrasonic sensors, and cell phones are examples of mobile sensors.
- *point sensor* if it is mounted at a fixed location along the roadway and observes traffic only at this particular location. Inductive-loop detectors, VIPS, pneumatic tubes, and RFID technology (e.g., transponder-reader system) are examples of point sensors.
- *space sensor* if it is up in the air and is able to take a snapshot of traffic on a stretch of road. Helicopters and satellites are examples of space sensors.

According to the extent to which a sensor intrudes into the roadway and traffic, the sensor can be

- *intrusive* if installation of the sensing system requires pavement work and interruption of traffic. Inductive-loop detectors and pneumatic tubes are examples of intrusive sensors.
- *nonintrusive* if installation of the sensing system does not require pavement work and interruption of traffic. VIPS and RFID technology are example of intrusive sensors.
- *off-roadway* if the sensor is not fixed to a location on the roadway—that is, the sensor can move with vehicles or float in the sky. GPS receivers, acoustic/ultrasonic sensors, cell phones, helicopters, and satellites are examples of space sensors.

1.3 DATA SOURCES

As example products of traffic sensors, two sets of data are presented below—Georgia State Route 400 (GA400) data and Next Generation Simulation (NGSIM) data. These data sets will be used in later chapters.

1.3.1 GA400 Data

GA400 is a toll road in Atlanta (Georgia, USA). Part of the road is freeway by design and is monitored by the NaviGATOR system—the—Georgia Department of Transportation’s intelligent transportation system. NaviGATOR’s video detection system (VDS) is the primary source of real-time information about current travel conditions. Approximately 1645 VDS stations are installed approximately every third of a mile along most major interstate highways in the Atlanta Metropolitan Area. These VDS cameras provide

continuous speed and volume data to the traffic management center and allow the system to generate travel times for the changeable message signs. NaviGATOR also uses about 500 full-color closed-circuit television cameras, positioned about every 1 mile on most major interstate highways in Atlanta. The closed-circuit television cameras have tilt, pan, and zoom capabilities, and serve as traffic cameras sending real-time footage to the operators at the traffic management center for enhanced situational awareness. The information collected from these cameras allows the operators to confirm incident details, dispatch rescue units, and request appropriate emergency resources.

Figure 1.9 shows a real-time traffic map of NaviGATOR in the Atlanta Metropolitan Area. On this map, roadway links are color-coded to highlight the level of congestion. In addition, the locations of some of the video cameras and changeable message signs are labeled on the map. A sample image from a video camera on GA400 is illustrated in the top left corner of the figure.

The data collected by the automated surveillance systems on GA400 were archived every day in the form of a single compressed file. This archived file contains observations at each station during the day. Each data entry represents 20 s of aggregation of classified vehicle counts, time

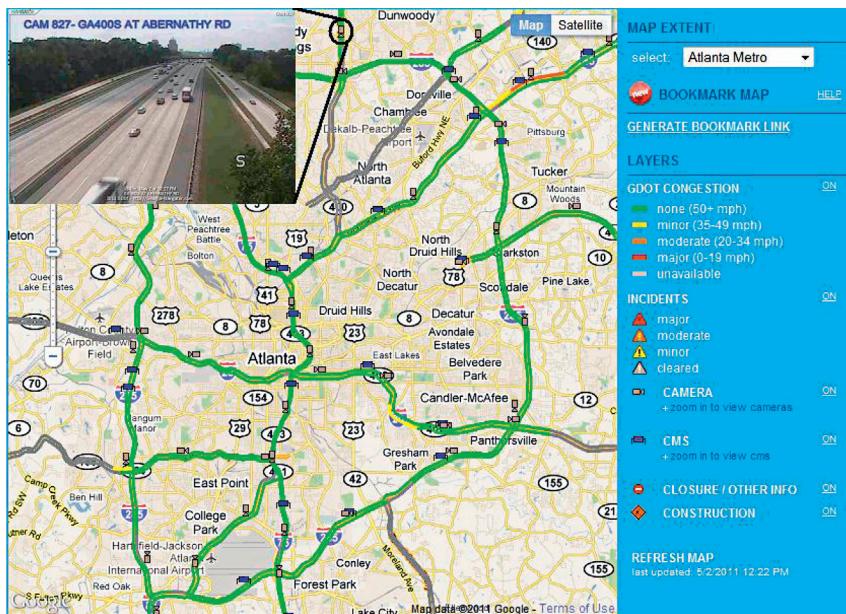


Figure 1.9 Georgia NaviGATOR. (Photo from <http://www.511ga.org/>)

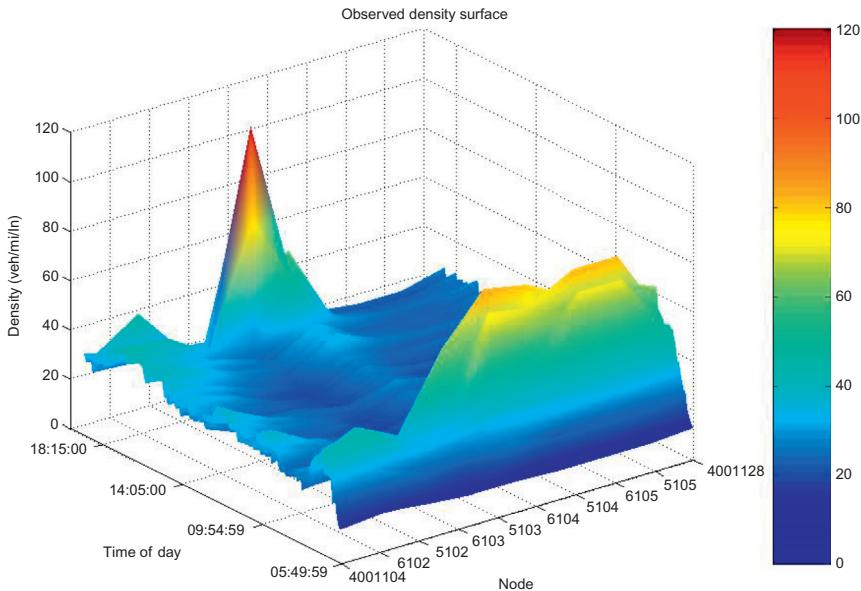


Figure 1.10 GA400 data sample.

mean speed, occupancy, etc. [Figure 1.10](#) illustrates three-dimensional traffic density (converted from field data collected on Friday, October 11, 2002) over time and space.

1.3.2 NGSIM Data

The NGSIM program was initiated by the Federal Highway Administration of the US Department of Transportation around 2000. The program developed a core of open behavioral algorithms in support of traffic simulation with a primary focus on microscopic modeling. To support the research and testing of the new algorithms, high-quality primary traffic and trajectory data were collected at multiple locations nationwide. The NGSIM program also actively engaged traffic simulation vendors to accelerate the inclusion of advanced or improved algorithms in the commercial models used across the world.

NGSIM vehicle trajectory data were collected for a set of sites including freeways, arterial roadways, and urban streets. [Figure 1.11](#) illustrates one of the sites on Interstate 80 in California. The left panel shows an aerial photo of the site where seven video cameras were set up on top of a 30-story building with each camera covering part of the study area. The right panel visualizes a camera and its perspective. These cameras shot the site

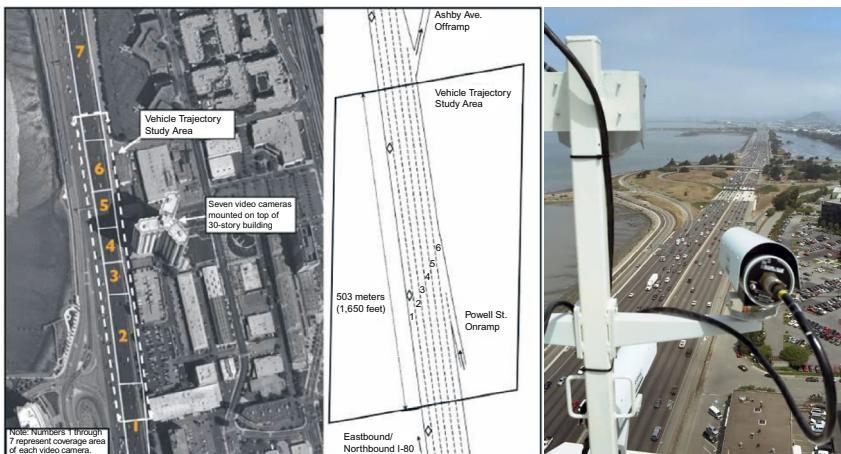


Figure 1.11 NGSIM data collection site. (*Photos from <http://www.fhwa.dot.gov/>*)



Figure 1.12 NGSIM data sample.

at different angles such that a vehicle entering from upstream is monitored continuously and consecutively by these cameras until it exits the study area.

Videos captured by these cameras were then processed by a customized software application which identifies, tracks, and records every vehicle's temporal-spatial positions as the vehicle traverses the study areas. The resultant vehicle trajectory data provided the precise location of each vehicle within the study areas every 0.1 s, resulting in detailed lane positions and locations relative to other vehicles. [Figure 1.12](#) illustrates a sample result of such vehicle trajectory data. The y -axis (not shown) is the highway running from south to north and the x -axis (not shown either) is time. Vehicle trajectories are so fine and dense that disturbances of traffic flow and its propagation are clearly visible like ripples in water.

PROBLEMS

1. Explain what an ITS is. What are the components of an ITS? Name a direction of recent development of ITS in the United States.
2. According to their working principle, traffic sensors can be classified into three types, what are they?

3. Provide an example of a mobile sensor. With a sketch, explain how it works, what kind of traffic data it is capable of collecting, and what advantages and disadvantages it has.
4. Provide an example of a point sensor. With a sketch, explain how it works, what kind of traffic data it is capable of collecting, and what advantages and disadvantages it has.
5. Provide an example of a space sensor. With a sketch, explain how it works, what kind of traffic data it is capable of collecting, and what advantages and disadvantages it has.

CHAPTER 2

Traffic Flow Characteristics I

According to their reporting mechanisms, traffic sensors can be classified into three categories: mobile sensors, point sensors, and space sensors. A mobile sensor resides in a vehicle, moves along with the vehicle, and logs the location of this particular vehicle over time. A point sensor sits at a fixed location on a roadway, sees the passage of vehicles above or under it, and reports traffic data only at this particular location over time. A space sensor flies in the sky, observes traffic on a stretch of road, and records the positions of vehicles at an instant of time over this particular stretch of road. It is interesting to see what traffic data reported by these sensors look like and, further, how traffic flow characteristics are determined from these data.

2.1 MOBILE SENSOR DATA

Let us start with mobile sensors. If a vehicle is equipped with a global positioning system (GPS) device, the device can report the vehicle's position as time progresses. Since GPS signals typically come once every second (i.e., at a frequency of 1 Hz), the GPS data may look similar to the data in [Table 2.1](#), where the vehicle's longitudinal x and lateral y displacements are relative to the vehicle's position at 09:00:00.

[Figure 2.1](#) shows the scenario in which a vehicle (with an on-board GPS device and ID number i) is moving on a roadway (drawn on the left) and the associated time-space diagram (drawn on the right). Every circle represents a GPS reading (only x is shown). If one connects these circles, the **trajectory** of this vehicle is obtained—that is, the location of the vehicle as a function of time: $x_i = x_i(t)$. It is easy to calculate the **speed** of the vehicle, \dot{x}_i , from the GPS data as illustrated in [Figure 2.1](#):

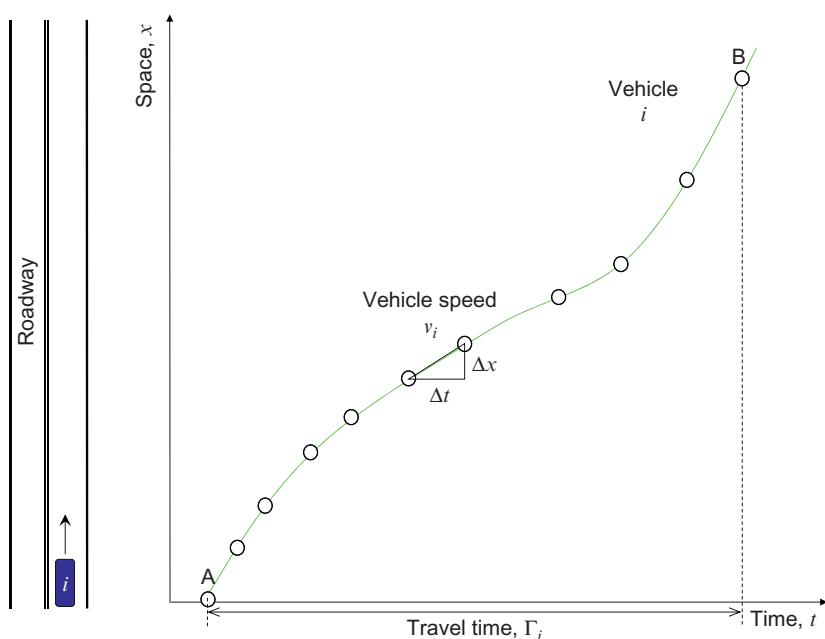
$$\dot{x}_i = \frac{\Delta x}{\Delta t}.$$

If the vehicle's trajectory is known and smooth, we can determine \dot{x}_i by taking the first derivative of the trajectory:

$$\dot{x}_i = \frac{dx}{dt}.$$

Table 2.1 GPS data

Time	x (feet)	y (feet)	x (m)	y (m)
09:00:00	0	0	0.0	0.0
09:00:01	3	0	0.9	0.0
09:00:02	5	0	1.5	0.0
09:00:03	7	0	2.1	0.0
09:00:04	10	1	3.0	0.3
09:00:05	15	4	4.6	1.2
09:00:06	18	9	5.5	2.7
09:00:07	21	12	6.4	3.7
09:00:08	23	12	7.0	3.7
09:00:09	27	12	8.2	3.7
09:00:10	30	12	30	12

**Figure 2.1** A trajectory.

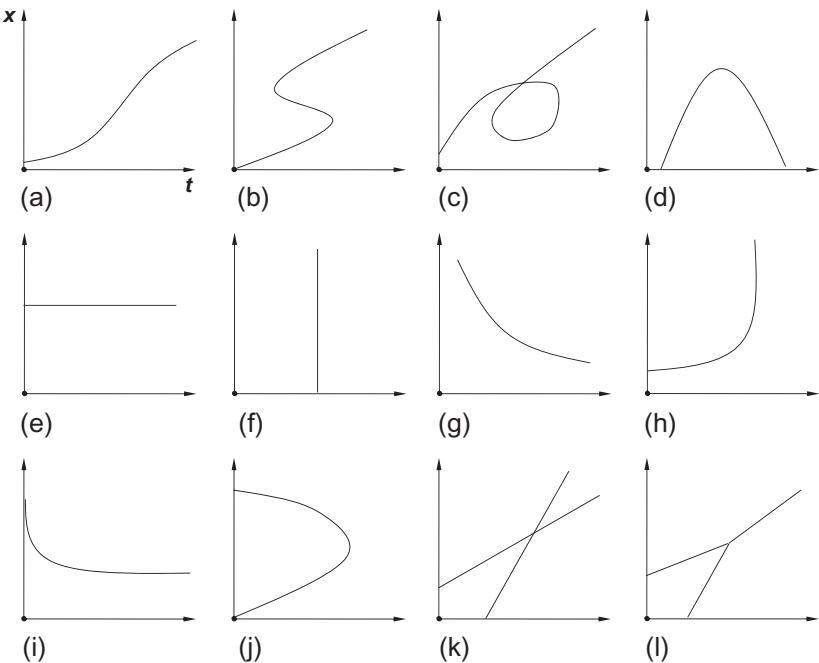


Figure 2.2 Vehicle trajectories.

The vehicle's **travel time**, t_i , between two points A and B can be directly read from the trajectory:

$$\Gamma_i = t_i^B - t_i^A.$$

Figure 2.2 illustrates some hypothetical vehicle trajectories, some of which are valid (i.e., trajectories that make sense), while some are not. Test yourself and see if you are able to identify which trajectories are valid and understand how these vehicles move. Plot (a) is valid, and it shows a vehicle moving in the positive x direction over time. Plot (b) is not a valid trajectory for the following reasons. If one draws a vertical line, it may intersect the trajectory several times. This means that at an instant of time the vehicle can appear at multiple locations simultaneously, which is impossible. For the same reason, plots (c) and (j) are not valid either. Plot (d) is valid, and the trajectory suggests that the vehicle first moves forward (i.e., in the positive x direction) and then, at some point in time, reverses. Plot (e) is valid, and simply suggests that the vehicle does not move (maybe parked). Plot (f) is impossible because it suggests an infinite speed (i.e., the tangent of the trajectory). Plot (g) is a valid since the vehicle just moves backward at a time-varying

speed. Plot (h) is likely but very unusual because the vehicle first moves at reasonable speeds and then almost flies at the end. Plot (i) is valid, and the vehicle gradually comes to a stop. Plot (k) can be interpreted in two ways: one is a two-lane scenario where a fast vehicle overtakes a slow vehicle; the other is a one-lane scenario where the fast vehicle collides with the slow vehicle and they exchange momentum. Plot (l) suggests that a fast vehicle catches up with a slow vehicle and then they move as a single unit thereafter.

2.2 POINT SENSOR DATA

If a point sensor (such as a loop detector or a video camera) is installed on the road at location x , this sensor will be able to observe vehicles passing above or under it. In a time-space diagram as illustrated in Figure 2.3, each vehicle will be counted (e.g., the tick marks) at this location. For example, during an observation period T , a total of N vehicles are counted by the sensor. N is referred to as the **traffic count**, which can be converted to the hourly equivalent rate of flow (referred as “**flow**” q hereafter) as follows:

$$q = \frac{N}{T}.$$

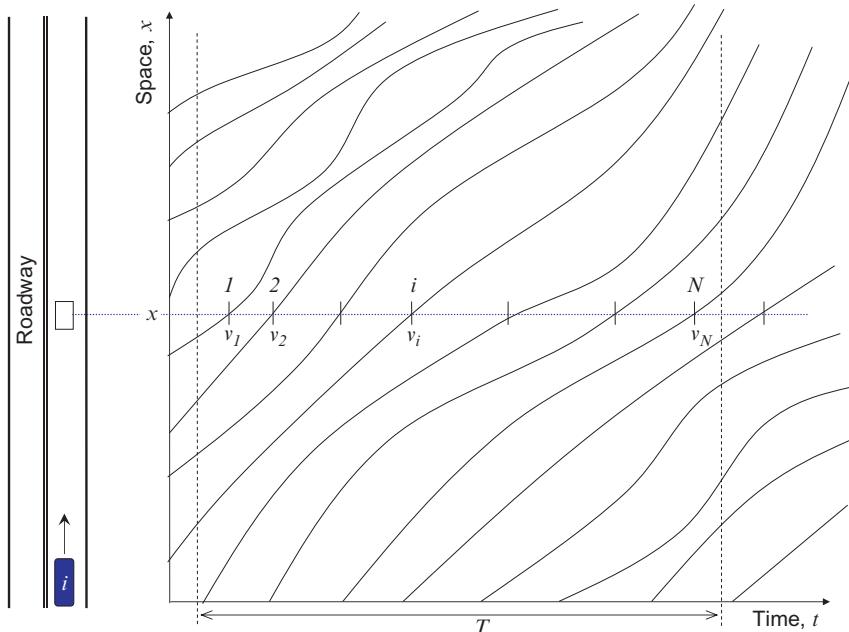


Figure 2.3 Point sensor data.

Headway h_i is defined as the temporal separation between two consecutive vehicles, and can be determined as:

$$h_i = t_i - t_{i-1}.$$

If one ignores the error due to incomplete headways of the first and last vehicles, the observation duration T can be expressed as

$$T = \sum_{i=1}^N h_i.$$

Both vehicles and point sensors have physical dimensions. If the sizes of the vehicles and sensors are taken into consideration, more information can be obtained from the time-space diagram (see Figure 2.4).

When a vehicle's front bumper enters the detection zone of a loop detector, a detection signal will be generated in the detector according to electromagnetism. When the vehicle's rear bumper exits the detection zone, the signal will drop. For an illustration of this effect, see the lower plot above

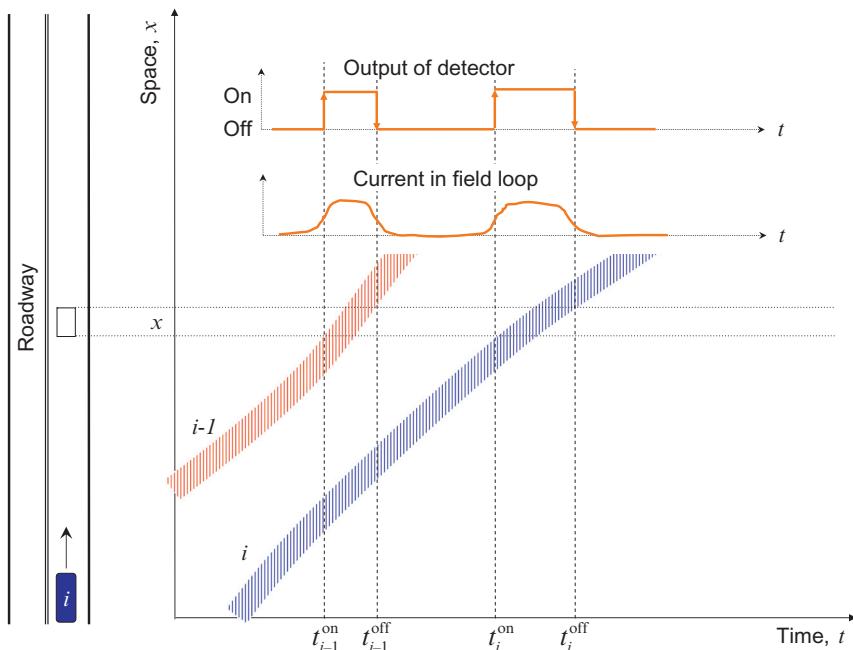


Figure 2.4 Loop detector data.

the two trajectories in [Figure 2.4](#). If a threshold is set properly, the loop detector outputs two states: “on” when a vehicle is above the loop and “off” when the loop detects no vehicle over it. When the loop outputs “on,” the loop is said to be “busy.” With the above setup, let us revisit some of the traffic flow characteristics discussed above and determine more characteristics:

Traffic count N : Since the on state consists of an upward transition and a downward transition of the detector output, one need only count either the upward transition or the downward transition consistently over all vehicles in order to obtain the traffic count.

Headway h_i : If one chooses reference points on all vehicles consistently (e.g., front bumpers), the headway between vehicles $i - 1$ and i can be calculated as $h_i = t_i^{\text{on}} - t_{i-1}^{\text{on}}$ and the time gap between them is $t_i^{\text{on}} - t_{i-1}^{\text{off}}$.

On time ξ_i : The duration from the moment when a vehicle’s front bumper enters the detection zone to the moment when the vehicle’s rear bumper exits the detection zone: $\xi_i = t_i^{\text{off}} - t_i^{\text{on}}$.

Vehicle speed \dot{x}_i : During the on time, vehicle i travels a distance of $d + l_i$ where d is the width of the loop (typically 6 feet or 1.8 m for small loops) and l_i is the length of the vehicle. Hence, the vehicle’s instantaneous speed can be determined as

$$\dot{x}_i = \frac{d + l_i}{\xi_i} = \frac{d + l_i}{t_i^{\text{off}} - t_i^{\text{on}}}.$$

Occupancy σ : In traffic flow theory, occupancy is defined as the percentage of time when a loop is busy—that is, when the loop detects vehicles above it. Hence, if the observation period is T , during which N vehicles are detected, the total on time is $\sum_{i=1}^N \xi_i$ and the occupancy is determined as

$$\sigma = \frac{\sum_{i=1}^N \xi_i}{T}.$$

Time-mean speed v_t : If one averages vehicle speeds observed at a point of roadway, one obtains a mean speed in the time domain, and hence such a mean speed is termed “time-mean speed.”

$$v_t = \frac{1}{N} \sum_{i=1}^N \dot{x}_i \dots \text{in the time domain.}$$

Interested readers are referred to [1], where there is detailed discussion of how various traffic flow characteristics are measured and calculated as well as how errors inherent in point sensors are introduced.

2.3 SPACE SENSOR DATA

If one takes aerial photos of a roadway from a helicopter, one is able to locate vehicles in each of these snapshots. For example, Figure 2.5 illustrates a snapshot taken at time t where vehicles are labeled as triangles. Some space-related traffic flow characteristics can be determined from these aerial photos:

Spacing s_i is defined as the spatial separation between two consecutive vehicles and can be determined as

$$s_i = x_{i-1} - x_i.$$

Density k is defined as the number of vehicles observed on a unit length of road and can be determined as

$$k = \frac{N}{L},$$

where L is the length of the stretch of road under observation and N is number of vehicles observed on this stretch of road. If one ignores the

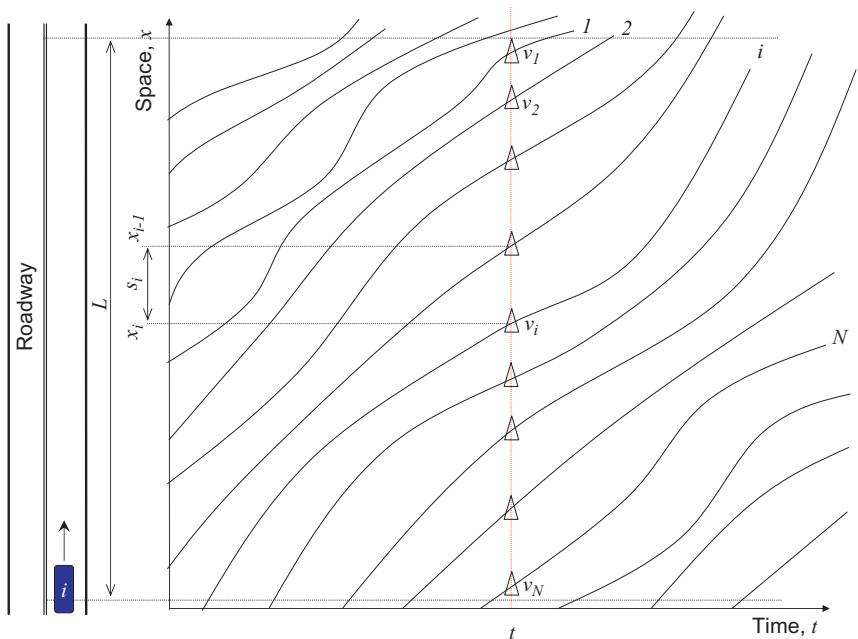


Figure 2.5 A snapshot of roadway.

error due to incomplete spacings of the first and last vehicles, the length of roadway L can be expressed as

$$L = \sum_{i=1}^N s_i.$$

Unfortunately, one is unable to determine the **vehicle speed** v_i from a single snapshot, but with two snapshots (at t_1 and t_2 , respectively), one is able to compare vehicle locations and find the distance traversed by each vehicle—that is, $\Delta x_i = x_i(t_2) - x_i(t_1)$. Since the time between the two snapshots $\Delta t = t_2 - t_1$ is known, the speed of each vehicle can be determined accordingly:

$$v_i(t) = \frac{\Delta x_i}{\Delta t}.$$

Space-mean speed v_s : If one averages vehicle speeds obtained from aerial photos, a mean speed in the space domain results, and hence such a mean speed is termed the “space-mean speed.”

$$v_s = \frac{1}{N} \sum_{i=1}^N \dot{x}_i \dots \text{in the space domain.}$$

2.4 TIME-SPACE DIAGRAM AND CHARACTERISTICS

The discussion so far has covered the three types of sensors (mobile, point, and space sensors), data reported by these sensors, and traffic flow characteristics determined with use of these data. It is informative to put everything together and form a complete picture. [Figure 2.6](#) shows a time-space diagram with vehicle trajectories where data reported by the three types of sensors are illustrated.

[Table 2.2](#) relates traffic flow characteristics to sensor types. Three categories of traffic flow characteristics are presented: flux, speed, and concentration. The characteristics are considered at two levels of detail: *microscopic* characteristics are vehicle specific and hence all bear subscript i , and *macroscopic* characteristics are aggregated measures and the aggregation can be done over vehicles, time, or space.

2.5 RELATIONSHIPS AMONG CHARACTERISTICS

So far, traffic flow characteristics have been introduced with the aid of field observations. It is interesting to investigate further the relationships among these traffic flow characteristics.

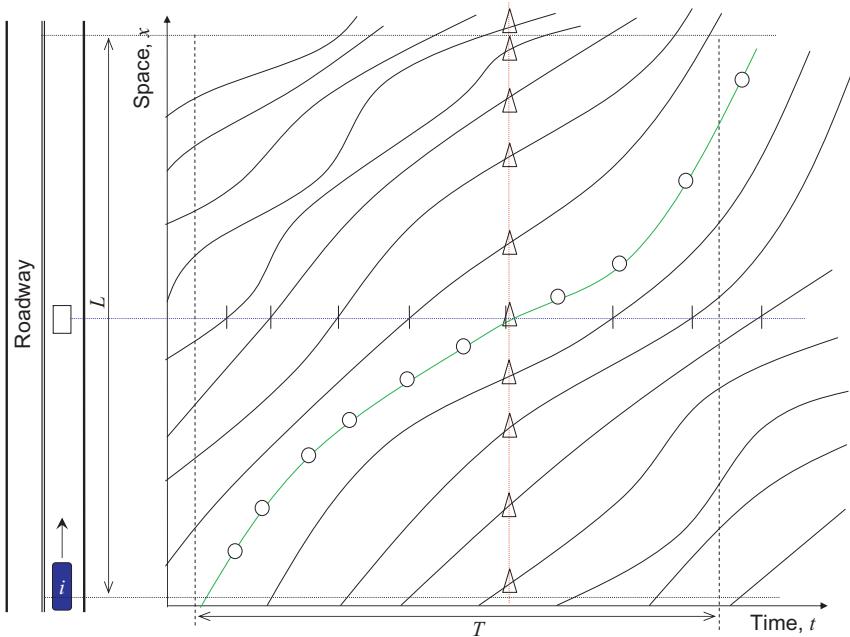


Figure 2.6 Time-space diagram and three types of sensors.

Table 2.2 Sensors and traffic flow characteristics

Category	Sensors	Microscopic characteristics	Macroscopic characteristics
Flux	Mobile	—	—
	Point	h_i	N, q
	Space	—	—
Speed	Mobile	\dot{x}_i	—
	Point	\dot{x}_i	v_t
	Space	\dot{x}_i	v_s
Concentration	Mobile	—	—
	Point	ξ_i	o
	Space	s_i	N, k

2.5.1 Flow, Speed, and Density

By definition, the following relationship holds as an *identity*:

$$q = k \times v_s;$$

that is, flow q is the product of density k and space-mean speed v_s .

2.5.2 Flow and Headway

From the above discussion, it follows that

$$q = \frac{N}{T},$$

$$T = \sum_{i=1}^n h_i,$$

$$q = \frac{N}{\sum_{i=1}^n h_i} = \frac{1}{\frac{1}{N} \sum_{i=1}^n h_i} = \frac{1}{h}.$$

Therefore, flow q is the reciprocal of average headway h . For example, a flow of 1200 vehicles per hour suggests an average headway of

$$\frac{1}{1200 \text{ vehicles per hour}} = \frac{3600 \text{ s/h}}{1200 \text{ vehicles per hour}} = 3 \text{ s.}$$

2.5.3 Density and Spacing

Similarly,

$$k = \frac{N}{L},$$

$$L = \sum_{i=1}^n s_i,$$

$$k = \frac{N}{\sum_{i=1}^n s_i} = \frac{1}{\frac{1}{N} \sum_{i=1}^n s_i} = \frac{1}{s}.$$

Therefore, density k is the reciprocal of average spacing s . For example, a density of 40 vehicles per mile (or 25 vehicles per kilometer) suggests an average spacing of

$$\frac{1}{40 \text{ vehicles per mile}} = \frac{5280 \text{ feet per mile}}{40 \text{ vehicles per mile}} = 132 \text{ feet}$$

or

$$\frac{1}{25 \text{ vehicles per kilometer}} = \frac{1000 \text{ m/km}}{25 \text{ vehicles per kilometer}} = 40 \text{ m.}$$

2.5.4 Time-Mean Speed and Space-Mean Speed

As discussed before, time-mean speed is vehicle speed averaged in the time domain, whereas space-mean speed is vehicle speed averaged in the space domain. Below we give an example that illustrates the difference between the two mean traffic speeds. Consider two lanes of traffic which is perfectly controlled so that there are only two streams of traffic: fast vehicles all travel at 60 miles per hour (or 96 km/h) in the inner lane and slow vehicles all move at 30 miles per hour (or 48 km/h) in the outer lane. Traffic flow in each lane is 1200 vehicles per hour, and lane change is prohibited. What is the time-mean speed and space-mean speed of traffic in both lanes?

Calculation of space-mean speed is straightforward, one simply averages the speed of vehicles observed on the road (see [Figure 2.7](#)). In 1 mile of the road, one observes a total of 60 vehicles, of which 20 vehicles are in the inner lane (1200 vehicles per hour/60 miles per hour) and 40 vehicles in the outer lane (1200 vehicles per hour/30 miles per hour). Therefore, space-mean speed is determined as

$$v_s = \frac{20 \times 60 \text{ mi/h} + 40 \times 30 \text{ mi/h}}{60} = 40 \text{ mi/h.}$$

Or in 1 km of the road, one observes a total of 37.5 vehicles, of which 12.5 vehicles are in the inner lane (1200 vehicles per hour/96 km/h) and 25 vehicles in the outer lane (1200 vehicles per hour/48 km/h). Therefore, space-mean speed is determined as

$$v_s = \frac{12.5 \times 96 \text{ km/h} + 25 \times 48 \text{ km/h}}{37.5} = 64 \text{ km/h.}$$

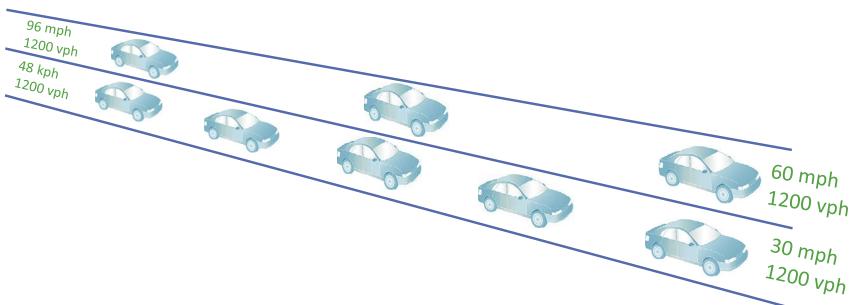


Figure 2.7 Time-mean speed versus space-mean speed. mph, miles per hour; vph, vehicles per hour.

For time-mean speed, one has to imagine a hypothetical observer standing at the roadside watching vehicles passing in front of him or her. As a result, the observer records 2400 vehicles in 1 h, of which 1200 vehicles are in the inner lane and 1200 vehicles are in the outer lane. Hence, by definition, time-mean speed is

$$\nu_t = \frac{1200 \times 60 \text{ mi/h} + 1200 \times 30 \text{ mi/h}}{2400} = 45 \text{ mi/h},$$

or

$$\nu_t = \frac{1200 \times 96 \text{ km/h} + 1200 \times 48 \text{ km/h}}{2400} = 72 \text{ km/h}.$$

Obviously, the results show that the two mean speeds are not equal. Wardrop [2] demonstrated that the following relationship between time-mean speed and space-mean speed always holds:

$$\nu_t = \nu_s + \frac{\sigma^2}{\nu_s},$$

where σ is the variance of vehicle speeds. It can be seen that time-mean speed ν_t is always greater than or equal to space-mean speed ν_s and they are equal only if the traffic is uniform—that is, all vehicles are traveling at the same speed ($\sigma = 0$).

Note that, in the above example, fast vehicles are overrepresented in the time-mean speed, with a fast to slow ratio of 1:1, while in reality the correct ratio is 1:2, which is the case in the calculation of space-mean speed. It can be further demonstrated that the space-mean speed is an unbiased estimate of the true traffic mean speed, while the time-mean speed is not.

2.5.5 Occupancy and Density

The following is reproduced from Ref. [3, Chapter 2]:

$$\begin{aligned} o &= \frac{1}{T} \sum_{i=1}^N \tau_i = \frac{1}{T} \sum_{i=1}^N \frac{d + l_i}{\dot{x}_i} \approx \frac{d + l}{T} \sum_{i=1}^N \frac{1}{\dot{x}_i} (\text{assume } l_i \rightarrow l) \\ &= (d + l) \frac{1}{T} \sum_{i=1}^N \frac{1}{\dot{x}_i} = (d + l) \left(\frac{N}{T} \right) \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{\dot{x}_i} \right) = (d + l) q \frac{1}{\nu_s} \\ &= (d + l) k = c_k. \end{aligned}$$

The approximately equal sign is based on the assumption of uniform vehicle length, $l_i = l$. With such an assumption, occupancy σ is proportional to density k , and the proportion coefficient c_k is the sum of loop width d and uniform vehicle length l .

2.6 DESIRED TRAFFIC FLOW CHARACTERISTICS

Control and optimization of traffic operations rely on an accurate understanding of traffic flow conditions, which in turn comes from field data collection.

Though the three types of sensors have their relative merits in terms of traffic data collection, they are practically very different, especially in terms of large-scale applications on a regular basis. Mobile sensors are not practical because not every vehicle is equipped with a GPS device. Though some vehicles may have GPS navigation systems or GPS-enabled cell phones, they are generally not intended for logging vehicle trajectories. Even if every vehicle had a GPS device and it were turned on to log vehicle trajectories, it would be prohibitive to make every driver comply with data extraction, let alone processing the data to generate a time-space diagram like that in [Figure 2.6](#). Space sensors are not suitable for applications on a regular basis. Think about the cost of hiring a helicopter flying over a road to observe traffic 24 hours a day and 7 days a week, not to mention the complexity of and time spent extracting traffic data from the huge number of aerial photos. Therefore, the only type of sensor that is feasible for automatic, regular, and large-scale applications is a point sensor such as a loop detector or a video camera (see Chapter 1 for details).

Traffic flow characteristics are not equally attractive when traffic control and management is concerned. For example, space-mean speed is preferred over time-mean speed as an unbiased estimate of the true mean traffic speed. In addition, space-mean speed is required in the identity $q = k \times v_s$ to calculate density or flow. Density is preferred over occupancy as a measure of traffic concentration. For example, the *Highway Capacity Manual* uses density as the measure of effectiveness to determine the level of service on freeways and multilane highways.

Hence, we have the following dilemma. On the one hand, space-based traffic flow characteristics such as space-mean speed and density are preferred. Therefore, space sensors are called for to provide measures of these traffic flow characteristics. On the other hand, space sensors are prohibitive to deploy on a large scale on a regular basis, while point sensors are

widespread (most intelligent transportation systems use point sensors), but report less attractive traffic flow characteristics such as time-mean speed and occupancy. Therefore, such a dilemma inevitably results in the estimation of space-based characteristics from point sensor data.

2.6.1 Determining Space-Mean Speed from Point Sensor Data

If individual vehicle speeds ($\dot{x}_i, i = 1, 2, \dots, N$) are available from a point sensor, these speeds can be used to determine space-mean speed as follows:

$$v_s = \frac{1}{\frac{1}{N} \sum_{i=1}^N \frac{1}{\dot{x}_i}}.$$

This mean is called the harmonic mean, in contrast to the arithmetic mean, which is the time-mean speed:

$$v_t = \frac{1}{N} \sum_{i=1}^N \dot{x}_i.$$

Unfortunately, many point sensor systems log only aggregated measures such as time-mean speed. In these systems, individual vehicle speeds are measured, but they are discarded after aggregation. As such, one has to resort to time-mean speed as a surrogate for space-mean speed, though one needs to recognize their difference, which might be considerable in some cases.

2.6.2 Determining Density from Point Sensor Data

Point sensor systems report occupancy, but not density. Using the above relationship between occupancy and density, one may be able to estimate density from occupancy:

$$k = \frac{o}{d + l},$$

though the reader must be cautioned about the implicit assumption of uniform vehicle length, which might be a strong one in some cases.

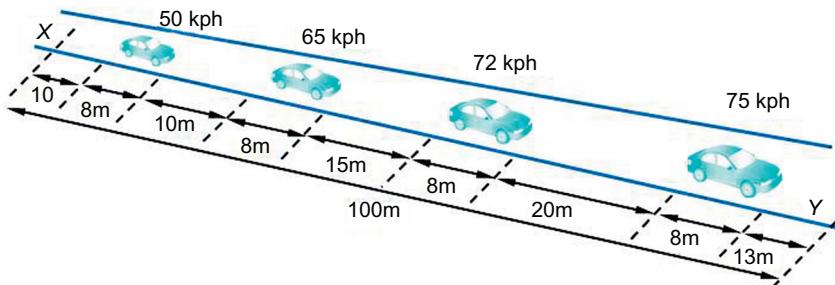
If a point sensor system time-stamps the passage of vehicles at two locations on the road with no vehicle appearing or exiting in between (e.g., a tunnel), it is possible to construct a curve showing the cumulative number of vehicles as a function of time at each location. Hence, density can be read directly from the cumulative curves. Interested readers are referred to [4] for further details.

Below are a few additional ways to calculate density k (not necessarily from point sensor data):

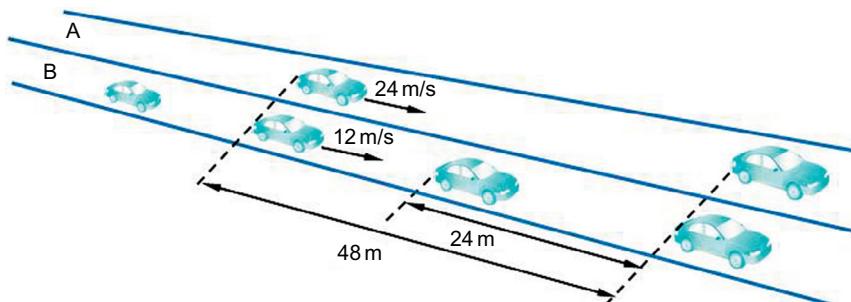
- $k = \frac{1}{s}$ if the average spacing s is known.
- $k = \frac{\bar{q}}{v_s}$. For point sensor data, replacing v_s with v_t sometimes yields more accurate k than that estimated from occupancy.
- Estimate density from travel times with the Kalman filter technique [5].

PROBLEMS

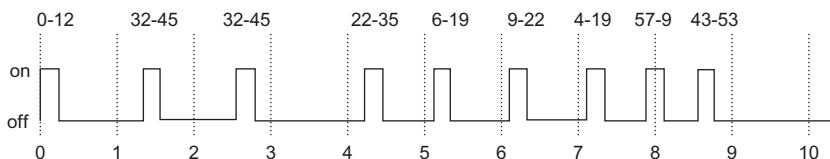
1. In a highway segment XY, an observer standing at location X counted four vehicles passing in front of him in 20 s. Their speeds are labeled in the figure. Find the flow, density, time-mean speed, and space-mean speed in this scenario.



2. The figure below illustrates two streams of uniform traffic in two lanes. In lane A, all vehicles are traveling at 24 m/s with a spacing of 48 m, while in lane B all vehicles are traveling at 12 m/s with a spacing of 24 m. Find the time-mean speed of all vehicles in the two lanes.

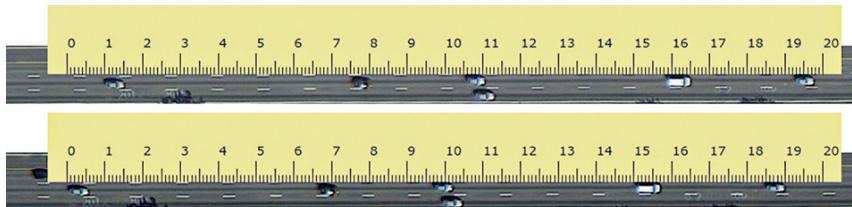


3. A traffic engineer counted vehicles on Route 9 and found that, on average, a vehicle passed in front of her at a rate of one every 5 s. If vehicles keep coming at this rate, what would be the equivalent hourly flow rate?
4. A roadside observer reports that five vehicles passed in the past 2 min. Their speeds are 30, 45, 20, 36, and 40 km/h respectively. Find the flow in vehicles per hour, time-mean speed, and space-mean speed. Which is greater, time-mean speed or space-mean speed?
5. For problem 2, find the space-mean speed.
6. Vehicle time headways and spacings were measured at a point along a highway, from a single lane, over the course of 1 h. The average values were calculated as 2.5 s per vehicle for the headway and 50 m per vehicle for the spacing. Calculate the average speed of the traffic.
7. A loop detector has recorded the information shown in the figure. The loop width is 6 feet. Each gate represents the brief period when a vehicle is in passing over the loop. “On” means a vehicle is in the detection zone, while “off” means no vehicle is in the detection zone. The numbers above each gate represent the duration of each gate. For example, “0-12” means a vehicle enters the detection zone at 0/60th second and exits at 12/60th second. The number of seconds is labeled at the bottom of the figure—for example, 1, 2, and 3 mean the first, second, and third seconds, respectively. Assume the vehicle length is uniformly 15 feet, and determine the following from the figure:



- a. Vehicle count during the observation period and the equivalent hourly flow rate
- b. Occupancy during the observation period
- c. Time-mean speed and space-mean speed
- d. Traffic density
- e. Estimate the speed from the speed-flow-density relationship and compare the result with that for (c). Is your estimated speed the time-mean speed or the space-mean speed?

8. The figure below shows aerial photos of a segment of Interstate 90. The two snapshots were taken 0.5 s apart and the scale of the ruler is 1:10 m. Using the section measured by the ruler and focusing on the middle lane traffic, find the vehicle spacings, traffic density, vehicle displacements, space-mean speed, and flow.



CHAPTER 3

Traffic Flow Characteristics II

3.1 GENERALIZED DEFINITION

In previous chapters, flow q and time mean speed ν_t were defined with the help of Figure 2.3 on the basis of point sensor data:

$$q = \frac{N}{T},$$
$$\nu_t = \frac{1}{N} \sum_{i=1}^N \dot{x}_i.$$

Similarly, density k and space mean speed ν_s were defined with the help of Figure 2.5 on the basis of space sensor data:

$$k = \frac{N}{L},$$
$$\nu_s = \frac{1}{N} \sum_{i=1}^N \dot{x}_i.$$

However, there is no common ground between the two sets of definitions, and the issue becomes more evident in Figure 2.6, where both cases are illustrated in the same figure. For example, the total number of vehicles N in the point sensor case is not necessarily the same as the total number N in the space sensor case. Similarly, vehicle speeds \dot{x}_i are not necessarily the same in both cases. The two sets of data are simply independent, though we adopted the same notation in both sets of definitions. As such, one is unable to conclude that the identity

$$q = k \times \nu_s$$

is guaranteed by definition. Therefore, the key to addressing this issue is to provide a common ground such that both sets of definition can be related in a single setting. For the convenience of further discussion, the above definition of flow, density, and mean speeds is referred to as *the Highway Capacity Manual (HCM) definition* hereafter since the definition is formally given in the HCM.

To find the common ground, let us rearrange the definition above as follows:

$$q = \frac{N}{T} = \frac{N \times dx}{T \times dx},$$

where dx denotes an infinitesimal distance (see Figure 3.1). If one ignores the slight error introduced by (possibly) incomplete trajectories of the first and last few vehicles, the physical meaning of the numerator is the sum of the distances traversed by all vehicles in area A during time period T :

$$d(A) = N \times dx = \sum_{i=1}^N \Delta x_i.$$

The denominator simply means the area of the time-space rectangle A bounded by T and dx , $|A|$. Hence, the definition of q can alternatively be expressed as the total distance traversed by all vehicles within A divided by the area of A :

$$q = \frac{d(A)}{|A|}.$$

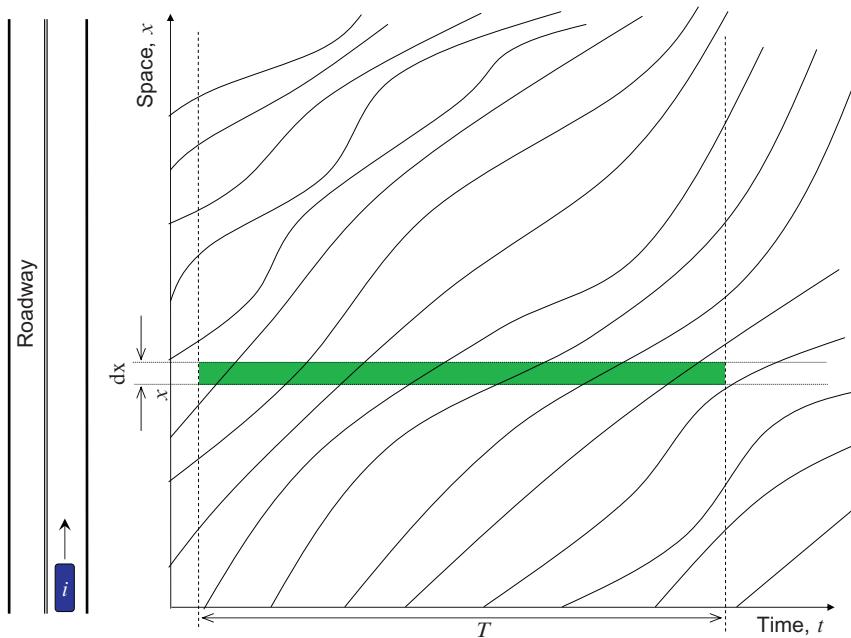


Figure 3.1 Time-space diagram with infinitesimal distance.

By definition in the HCM, the mean speed of vehicles, ν , is the total distance traveled by all vehicles divided by the total travel time of these vehicles. The total distance traveled by all vehicles within rectangle A is $d(A) = N \times dx$. The total time spent by all vehicles within A is

$$t(A) = \sum_{i=1}^N \frac{dx}{\dot{x}_i}.$$

Therefore,

$$\nu = \frac{d(A)}{t(A)} = \frac{N \times dx}{\sum_{i=1}^N \frac{dx}{\dot{x}_i}} = \frac{N \times dx}{dx \times \sum_{i=1}^N \frac{1}{\dot{x}_i}} = \frac{1}{\frac{1}{N} \sum_{i=1}^N \frac{1}{\dot{x}_i}}.$$

This is the harmonic mean, which corresponds to the space mean speed presented in the point sensor scenario.

Similarly, density k can be represented as

$$k = \frac{N}{L} = \frac{N \times dt}{L \times dt},$$

where dt denotes an infinitesimal duration (see [Figure 3.2](#)). Following the same argument as above, L and dt define a time-space rectangle A. The numerator is the sum of the times spent by all vehicles within A, $t(A)$, and the denominator is the area of the rectangle, $|A|$:

$$t(A) = N \times dt = \sum_{i=1}^N dt_i,$$

$$|A| = L \times dt.$$

Hence,

$$k = \frac{t(A)}{|A|}.$$

The total distance traveled by all vehicles within A is $d(A) = \sum_{i=1}^N dt \times \dot{x}_i$. Hence, the mean speed of these vehicles is

$$\nu = \frac{\sum_{i=1}^N dt \times \dot{x}_i}{N \times dt} = \frac{1}{N} \sum_{i=1}^N \dot{x}_i.$$

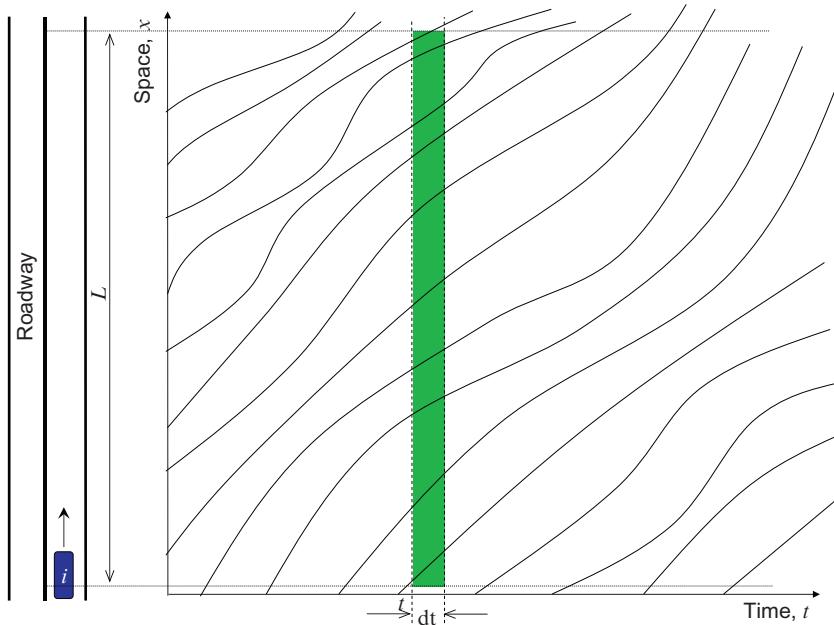


Figure 3.2 Time-space diagram with infinitesimal duration.

This is the arithmetic mean, which corresponds to the space mean speed determined in the space sensor scenario.

The above discussion suggests that a time-space rectangle may serve as the common ground to unify the definition of flow q , mean speed v , and density k . [Figure 3.3](#) illustrates a general time-space rectangle A covering length L (bounded by upstream location x_{lo} and downstream location x_{hi}) and duration T (bounded by instants t_{lo} and t_{hi}). On the basis of A , the three traffic flow characteristics can be defined as follows:

$$q(A) = \frac{d(A)}{|A|},$$

$$k(A) = \frac{t(A)}{|A|},$$

$$v(A) = \frac{d(A)}{t(A)}.$$

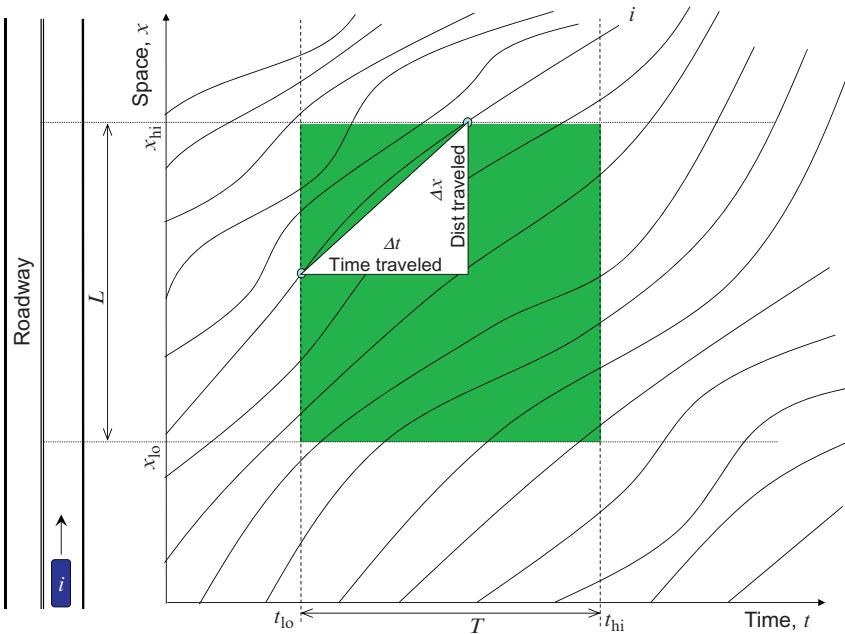


Figure 3.3 Time-space diagram with rectangle.

Therefore, the identity $q = k \times \nu$ is now guaranteed by definition. The question that remains is how to determine $d(A)$, $t(A)$, and $|A|$. Take an arbitrary vehicle i . For example, the vehicle enters A at location $x_i(t_{lo})$ or x_{lo} whichever comes later and at time t_{lo} or $t_i(x_{lo})$ whichever comes later; the vehicle exits A at location x_{hi} or $x_i(t_{hi})$ whichever comes earlier and at time t_{hi} or $t_i(x_{hi})$ whichever comes earlier. Hence, the distance traveled by vehicle i in A is

$$\Delta x_i = \min(x_{hi}, x_i(t_{hi})) - \max(x_i(t_{lo}), x_{lo}).$$

Therefore, the total distance traveled by all vehicles in A is

$$d(A) = \sum_{i=1}^N \Delta x_i.$$

The time spent by vehicle i in A is

$$\Delta t_i = \min(t_{hi}, t_i(x_{hi})) - \max(t_{lo}, t_i(x_{lo})).$$

Hence, the total time spent by all vehicles in A is

$$t(A) = \sum_{i=1}^N \Delta t_i.$$

The area of A is simply

$$|A| = L \times T.$$

Therefore, all the quantities needed to calculate the flow, mean speed, and density have been determined.

The following question naturally arises: Does the common ground have to be a rectangle? The answer is no. Actually, any time-space region will work as long as the region is closed (see [Figure 3.4](#)). The above definition was originally proposed by Edie [6]. Readers are referred to the original paper for an in-depth discussion. For convenience, the above set of definitions of flow, mean speed, and density based on a time-space region is referred to as **the generalized definition**.

It can be seen that the HCM definition is a special case of the generalized definition. For example, if one takes a time-space region like the one in [Figure 3.1](#) and allows $dx \rightarrow 0$, a point sensor scenario results, while one obtains the space sensor scenario in [Figure 3.2](#) if one keeps L constant and makes $dt \rightarrow 0$.

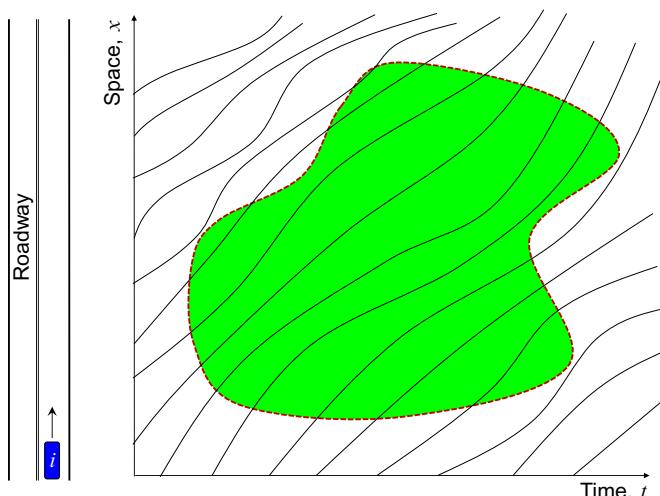


Figure 3.4 Time-space diagram with general region.

3.2 THREE-DIMENSIONAL REPRESENTATION OF TRAFFIC FLOW

The following discussion is based on Ref. [7]. Interested readers are referred to the original paper for in-depth information.

So far, we have been working with a time-space diagram and vehicle trajectories, on the basis of which a connection is made to traffic flow characteristics. A time-space diagram is a two-dimensional representation, and the discussion can be made more informative if we adopt a three-dimensional perspective. Taking the family of vehicle trajectories in Figure 3.3, for example, we see these trajectories lie on the same plane defined by time (t) and space (x). These vehicles are numbered cumulatively (i.e., ID = 1, 2, 3, ...) in the order they appear on the road, and each vehicle is elevated along the third dimension to the height corresponding to the vehicle's ID (i.e., vehicle 1 raised to height 1, vehicle 2 raised to 2, and so on). Let us call the third dimension the cumulative number of vehicles (N) and denote the surface that passes these elevated vehicle trajectories $N(x, t)$. Figure 3.5 illustrates two examples of such a three-dimensional representation adopted from Ref. [7, 8].^{1,2}

What makes this three-dimensional representation interesting is that it can be used to illustrate and relate some key concepts of traffic flow conveniently. For example, if one cuts $N(x, t)$ in the lower part of Figure 3.5 using plane $t = 20$, one obtains the shape PQN and its projection P'Q'N' on the $N - x$ plane. Curve P'N' can be interpreted as the snapshot taken at time $t = 20$, which shows the location of each vehicle at this moment. Figure 3.6 illustrates more examples of such curves (they look like stairs before smoothing), where $N(x, t)$ in the upper part of Figure 3.5 is cut at different instants and projected it onto the $N - x$ plane. Each curve represents a snapshot taken at the time instant indicated on that curve. For example, the lowest curve is a snapshot taken at time $t^{(1)}$. If one draws a horizontal line at height $N = 2$, the intersection of this line and the curve labeled $t^{(1)}$ is the location of vehicle with ID 2 at time $t^{(1)}$. Note that this line needs to be slightly lower—say, at height $N = 1.999$ —to avoid multiple intersections, and the same applies hereafter. Similarly, the intersection of line $N = 2$ and curve $t^{(2)}$ is the location of vehicle 2 at time $t^{(2)}$. The distance between

¹ $N(x, t)$ in the upper part of Figure 3.5 is not smoothed, while that in the lower part is smoothed. By default, a smoothed surface is assumed in order to take derivatives.

² If two trajectories intersect, the surface will be multivalued at a time-space point. Makigami et al. [7] showed how to resolve the problem.

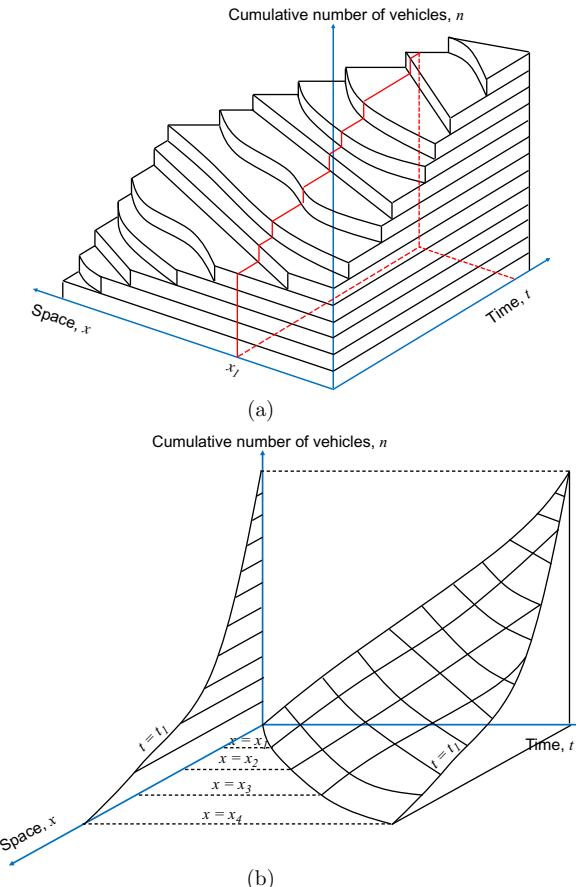


Figure 3.5 Three-dimensional representation examples.

the two intersections is the distance traversed by vehicle 2 from $t^{(1)}$ to $t^{(2)}$. If an $N-x$ curve at time t is smoothed (like curve $P'N'$ in Figure 3.5), the tangent of the curve denotes the **density** k at this instant. Note that the tangent slants down (because lower-numbered vehicles are in front), so it has a negative value. Hence,

$$k|_t = -\frac{dN}{dx}\Big|_t .$$

Similarly, if one cuts the three-dimensional model with a plane passing a specific location and parallel to the $N-t$ plane, one obtains a curve representing the cumulative number of vehicles passing this location over

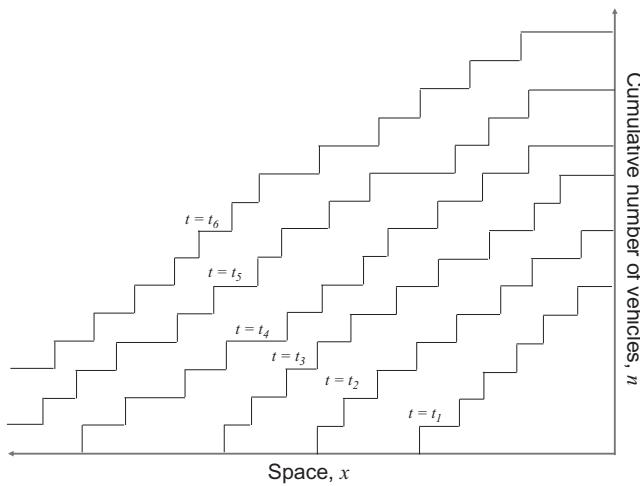


Figure 3.6 The N - x diagram.

time—for example, the curves in the lower part of [Figure 3.5](#), such as the $x = 0, 2, 4, 6$ curves, and the curves in [Figure 3.7](#). If one draws a horizontal line at height $N = 2$ in [Figure 3.7](#), the intersection of this line and the curve labeled $x^{(2)}$ indicates the time when the vehicle with ID 2 passes location $x^{(2)}$. Similarly, the intersection of line $N = 2$ and curve $x^{(3)}$ is the time when vehicle 2 passes location $x^{(3)}$. The distance between the two intersections is the travel time for vehicle 2 to traverse from location $x^{(2)}$ to location $x^{(3)}$. If an N - t curve at location x is smoothed, the tangent of this curve denotes the **flow** q at this location:

$$q|_x = \frac{dN}{dt}|_x.$$

Therefore, flow and density can be expressed as partial differentials of the surface $N(x, t)$:

$$q = \frac{\partial N(x, t)}{\partial t},$$

$$k = -\frac{\partial N(x, t)}{\partial x}.$$

In addition, if one projects a region on the surface $N(x, t)$ (e.g., region A in [Figure 3.8](#)) onto the x - t , N - t , and N - x planes, one obtains three projections— A_N , A_x , and A_t , respectively. Makigami et al. [7] demonstrated

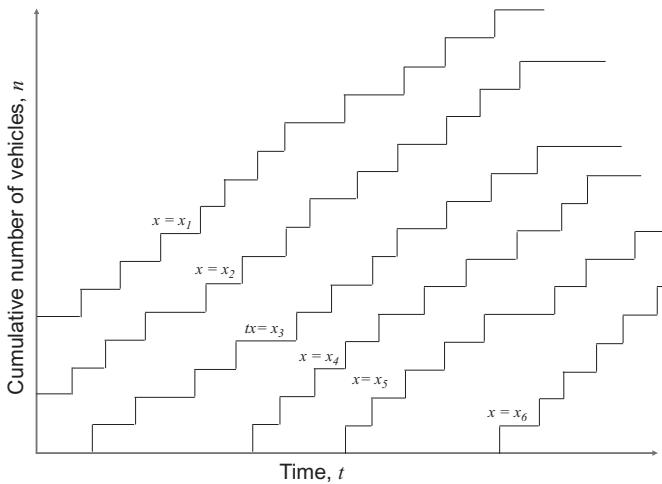


Figure 3.7 The N - t diagram.

that the following relationships hold:

$$A^2 = A_N^2 + A_x^2 + A_t^2,$$

$$q = \frac{A_t}{A_N},$$

$$k = \frac{A_x}{A_N},$$

$$\nu = \frac{A_t}{A_x}.$$

Figure 3.9 summarizes the previous graphics in one figure. Plot A shows vehicle trajectories in the x - t plane. Plot D raises the vehicle trajectories to their corresponding height and forms the three-dimensional surface $N(x, t)$. Plot B shows two N - t curves observed at locations $x = x_2$ and $x = x_4$. Plot C depicts two N - x curves resulting from snapshots taken at $t = t_6$ and $t = t_8$.

In addition to deepening the understanding of traffic flow and its characteristics, the three-dimensional model can be used to solve practical problems. For example, as mentioned before, space-based measures such as density and space mean speed are desired. In addition, determination of these traffic flow characteristics based on generalized definition (as opposed

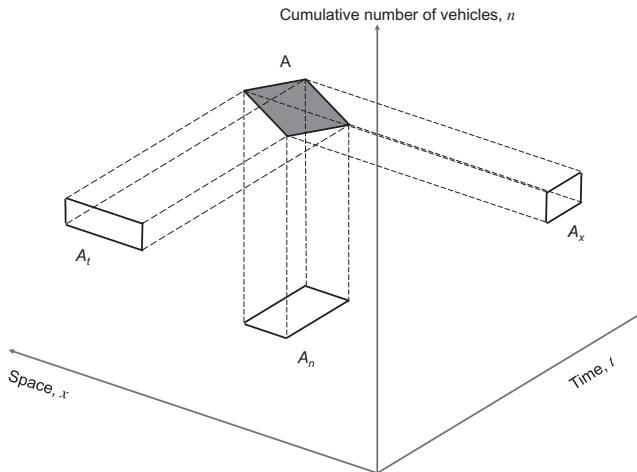


Figure 3.8 Projection of an $N-t-x$ region.

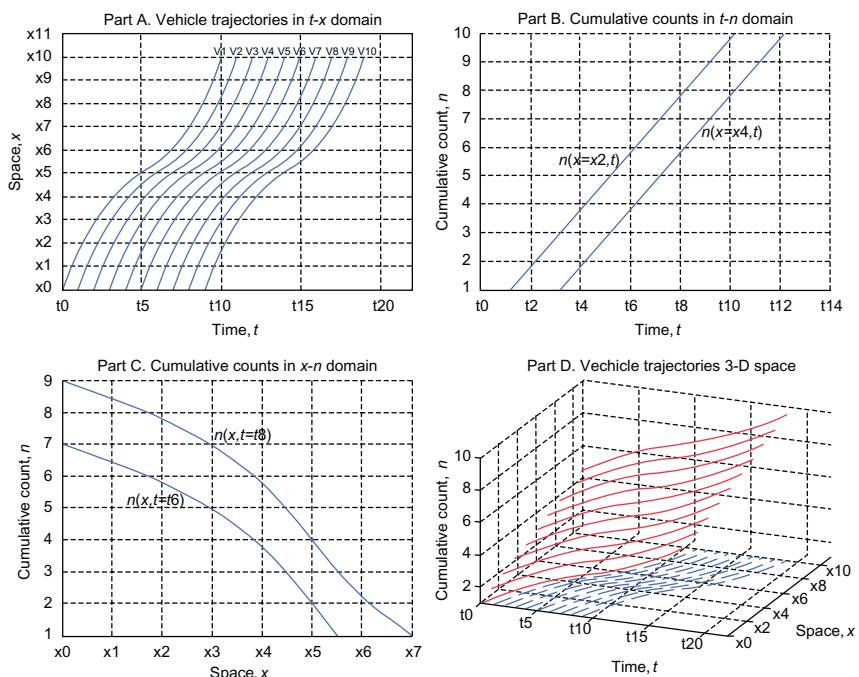
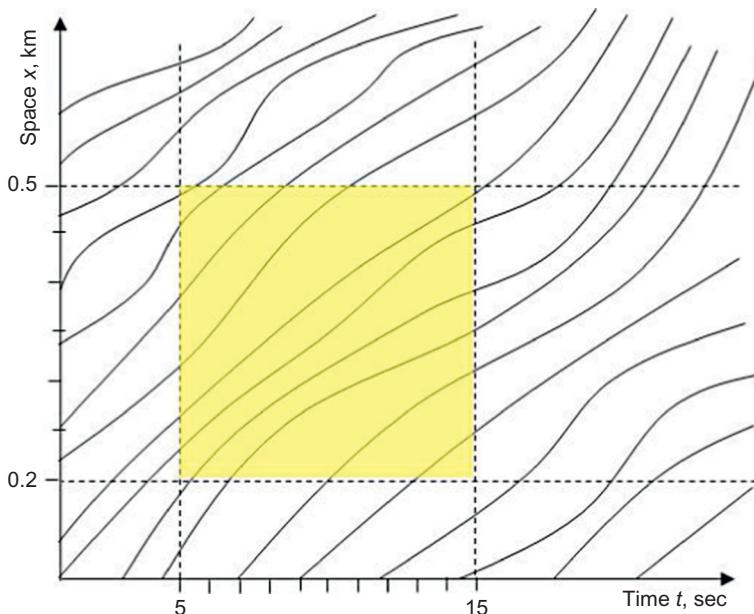


Figure 3.9 Three-dimensional representation of traffic flow.

to the HCM definition) is also preferred. However, the widely deployed intelligent transportation systems consist mostly of point sensors, which are generally unable to report space-based traffic flow characteristics. Interested readers are referred to [4] to learn how the three-dimensional representation helps address the problem by computing the desired traffic flow characteristics based on the preferred definition from intelligent transportation system data.

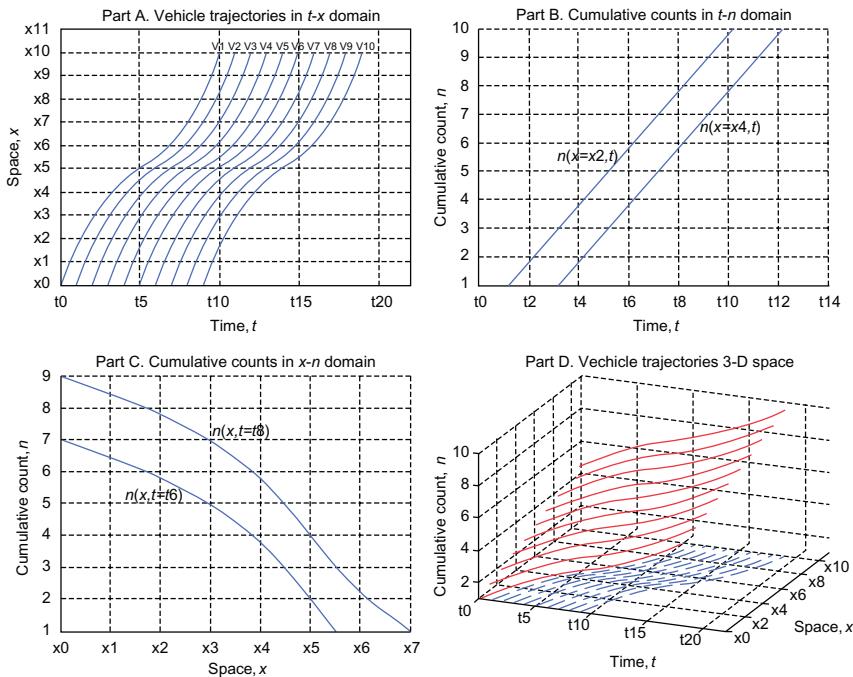
PROBLEMS

1. The figure below shows a set of vehicle trajectories in the time-space plane. An analysis of traffic flow characteristics is performed on the basis of a bounding box. As shaded in the figure, the box is confined in time between $t = 5$ s and $t = 15$ s and in space between $x = 0.2$ km and $x = 0.5$ km.



- a. Identify vehicles that traverse the bounding box.
- b. Transcribe the time and location when each of the above-mentioned vehicles enter and exit the bounding box.

- c. Find the duration and distance that each vehicle traveled in the bounding box.
 - d. Find the travel speed for each vehicle when it traveled in the bounding box.
 - e. Find the average speed of all vehicles that traveled in the bounding box.
 - f. With use of the generalized definition, find the flow, space mean speed, and density based on conditions in the bounding box.
 - g. State the general relationship among flow, speed, and density. Then verify the relationship using the result in (f). Does this relationship hold? Does it hold strictly? Is it guaranteed by definition?
 - h. Now, focus on observations made at location $x = 0.2$ km during time $t = 5$ s to $t = 15$ s. Find number of vehicles passing this location during this time period. Calculate the equivalent hourly flow rate.
 - i. Next, focus on observations made at instant $t = 5$ s on a road section between $x = 0.2$ km and $x = 0.5$ km. Find the number of vehicles within this section of road at that moment. Calculate the equivalent density.
 - j. With the average speed found in (e), flow in (h), and density in (i), another set of speed, flow, and density results. Redo (g) and comment on your findings.
2. Part D in the figure below illustrates a stream of traffic in a three-dimensional representation which is then projected onto three planes: the time-space plane (t - x plane), the cumulative number-time plane (n - t plane), and the cumulative number-space plane (n - x plane). Time is labeled in minutes—for example, “t5” means $t = 5$ min—and space is labeled in kilometers—for example “x2” means $x = 2$ km.
 - a. Redo the set of problems in 1 based on bounding box with time between $t = 5$ and $t = 10$ min and space between $x = 2$ km and $x = 4$ km, as shaded. When you answer question (h), assume the location is $x = 2$ km during time $t = 5$ to $t = 10$ min. When you answer (i), assume the instant is $t = 5$ min on the road section between $x = 2$ km and $x = 4$ km.
 - b. With use of part B in the figure, find the instantaneous flow observed at time $t = 6$ min and location $x = 2$ km.



- c. With use of part C in the figure, find the instantaneous density observed at time $t = 6$ min and location $x = 2$ km.
- d. With use of the space mean speed found in (a), flow found in (b), and density found in (c), verify the fundamental relationship among flow, space mean speed, and density. Comment on your result.

CHAPTER 4

Equilibrium Traffic Flow Models

In the previous chapter, it was shown that the following relationship holds among flow q , density k , and space mean speed ν (the subscript “ s ” is dropped unless it is necessary to distinguish space mean speed ν_s from time mean speed ν_t):

$$q = k \times \nu.$$

This relationship is an *identity* since it is self-guaranteed under the generalized definition. One may wish to know what other relationships exist among the three traffic flow characteristics. For example, is there any pairwise relationship between flow and density, density and speed, and speed and density? This chapter attempts to address these questions.

4.1 SINGLE-REGIME MODELS

Let us start with field observations. Figure 4.1 illustrates an image captured by a point sensor (a video camera in Georgia NaviGAtor, Georgia’s intelligent transportation system). The point sensor constitutes an observation station consisting of a group of imaginary detectors with one detector in each lane.

As discussed before, traffic data can be extracted from video images by means of image processing. Figure 4.2 shows a portion of a daily report from a video camera. Each row represents observations aggregated over 20 s over all lanes. Column A contains the station ID, column B contains the time stamp of each observation, column C contains the status of the detectors of this station (there are four lanes at this station and hence there are four detectors. “OK” means the corresponding detector is working properly, while “NO_ACT” means no actuation), columns E through H contain classified traffic counts (only column E is shown here because of limited space), column I contains the occupancy, column K contains the time mean speed, column M contains the average vehicle length, and column P contains the density (estimated by a proprietary recipe).

The point sensor data are plotted in Figure 4.3, where the top-left plot shows the speed-density relationship, the top-right plot shows the speed-flow relationship, the bottom-left plot shows the flow-density



Figure 4.1 An image captured by a point sensor. (From NaviGAtor.)

relationship, and the bottom-right plot shows the speed-spacing relationship. The “cloud” in Figure 4.3 represents 1 year’s worth of field observations aggregated to 5 min (i.e., each point in the figure represents the traffic condition observed in 5 min). The traffic speed here is the time mean speed since it is impossible to calculate the space mean speed from aggregated point sensor data. Density is estimated from flow and speed. The large dots represent the average of the “cloud.” The plots in Figure 4.3 were generated with use of traffic data collected at a fixed location. Therefore, such plots are location specific—that is, plots generated from different locations may differ. In addition, time information is lost in the figure—that is, one could not deduce the time when a data point was observed. As such, the figure actually depicts an *equilibrium* or *steady-state* relationship. Consequently, models of such a relationship without a reference to time are termed “equilibrium models” or “steady-state models.”

Noticeably, each plot in Figure 4.3 exhibits a trend which suggests a pairwise relationship among flow, speed, and density, though such a relationship is of statistical significance. For example, the top-left plot reveals a decreasing relationship between speed and density with two intercepts intuitively known. The intercept on the space x -axis represents a “Sunday morning” scenario where there are very few vehicles on the road (i.e., $k \rightarrow 0$). Hence, one may drive at the desired speed without being blocked by a slow driver ($v \rightarrow v_f$, the free-flow speed). The other intercept corresponds to a “Friday afternoon peak” scenario where everyone rushes home. As such, the road is jammed ($k \rightarrow k_j$, the jam density), resulting

A	B	C	E	I	K	M	P	Q
#station_id	sample_start	status	volume_auto	time_occupancy	time_speed	length	density	gap
803	4001134 2003-10-08 05:41:00 000 EDT	INO_ACTOKIOKOK	6	0.014	49.1503	14.7638	8.1	796.9
804	4001134 2003-10-08 05:41:20 000 EDT	OKIOKIOKOK	23	0.0276	47.4313	14.5451	13.873	383.8095
805	4001134 2003-10-08 05:41:40 000 EDT	OKIOKIOKOK	17	0.031	48.2745	14.5607	15.5	353.4524
806	4001134 2003-10-08 05:42:00 000 EDT	OKIOKIOKOK	20	0.0473	46.1478	14.7638	18.6429	292.9821
807	4001134 2003-10-08 05:42:20 000 EDT	OKIOKIOKOK_ACT	9	0.0136	53.0384	14.7638	6.7143	761.4286
808	4001134 2003-10-08 05:42:40 000 EDT	OKIOKIOKOK	14	0.02	48.8921	14.7638	11.1579	895.8947
809	4001134 2003-10-08 05:43:00 000 EDT	OKIOKIOKOK	8	0.0086	45.0776	14.7638	5.8782	1055.6882
810	4001134 2003-10-08 05:43:20 000 EDT	OKIOKIOKOK	13	0.0252	50.8922	13.8642	10.7742	538.0645
811	4001134 2003-10-08 05:43:40 000 EDT	OKIOKIOKOK	8	0.0221	52.8492	14.7638	11.2632	619.4211
812	4001134 2003-10-08 05:44:00 000 EDT	OKIOKIOKNO_ACT	10	0.0138	54.9913	14.7638	7.375	779
813	4001134 2003-10-08 05:44:20 000 EDT	OKIOKIOKOK	17	0.019	50.2191	14.7638	10.7	1558.74
814	4001134 2003-10-08 05:44:40 000 EDT	OKIOKIOKOK	14	0.0446	44.9162	14.7638	19.8	295.2571
815	4001134 2003-10-08 05:45:00 000 EDT	OKIOKIOKOK	24	0.0557	43.5066	17.0135	18.2857	295.1857
816	4001134 2003-10-08 05:45:20 000 EDT	OKIOKIOKOK	15	0.0372	49.8987	14.7638	17.6522	439.5217
817	4001134 2003-10-08 05:45:40 000 EDT	OKIOKIOKOK	8	0.0115	50.9524	14.7638	6.7	1667.55
818	4001134 2003-10-08 05:46:00 000 EDT	OKIOKIOKOK	18	0.0229	49.7929	14.7638	11.5556	461.1111
819	4001134 2003-10-08 05:46:20 000 EDT	INO_ACTOKIOKOK	8	0.0236	51.4406	14.7638	10.2143	544.3571
820	4001134 2003-10-08 05:46:40 000 EDT	OKIOKIOKOK	6	0.0092	46.555	14.7638	4.1538	1429.615
821	4001134 2003-10-08 05:47:00 000 EDT	OKIOKIOKOK	14	0.0239	51.5573	14.7638	12.1842	1225.237
822	4001134 2003-10-08 05:47:20 000 EDT	OKIOKIOKOK	24	0.04	45.6946	14.7638	21.2769	330.6923
823	4001134 2003-10-08 05:47:40 000 EDT	OKIOKIOKOK	11	0.0309	44.9755	19.685	11.7812	484.8125
824	4001134 2003-10-08 05:48:00 000 EDT	OKIOKIOKOK	6	0.0094	49.6365	14.7638	4.8824	1318
825	4001134 2003-10-08 05:48:20 000 EDT	OKIOKIOKOK	8	0.0125	49.4766	14.7638	4.875	1271
826	4001134 2003-10-08 05:48:40 000 EDT	OKIOKIOKOK	16	0.0495	47.9935	14.7638	23.0476	253.7381
827	4001134 2003-10-08 05:49:00 000 EDT	OKIOKIOKOK	26	0.0421	47.6679	14.7638	20.1429	578.4935
828	4001134 2003-10-08 05:49:20 000 EDT	OKIOKIOKOK	23	0.0577	49.0984	16.0946	20.3115	265.3279
829	4001134 2003-10-08 05:49:40 000 EDT	OKIOKIOKOK	20	0.0412	47.9673	14.7638	20.451	268.1961
830	4001134 2003-10-08 05:50:00 000 EDT	OKIOKIOKOK	14	0.0334	44.5611	14.3701	13.9714	464.8887
831	4001134 2003-10-08 05:50:20 000 EDT	OKIOKIOKOK	15	0.0303	48.8531	14.7638	13.7027	404.027
832	4001134 2003-10-08 05:50:40 000 EDT	OKIOKIOKOK	25	0.0488	45.2629	14.7638	24.7344	250.25
833	4001134 2003-10-08 05:51:00 000 EDT	OKIOKIOKOK	23	0.0549	44.1544	14.7638	27.6667	199.3881

Figure 4.2 Point sensor data.

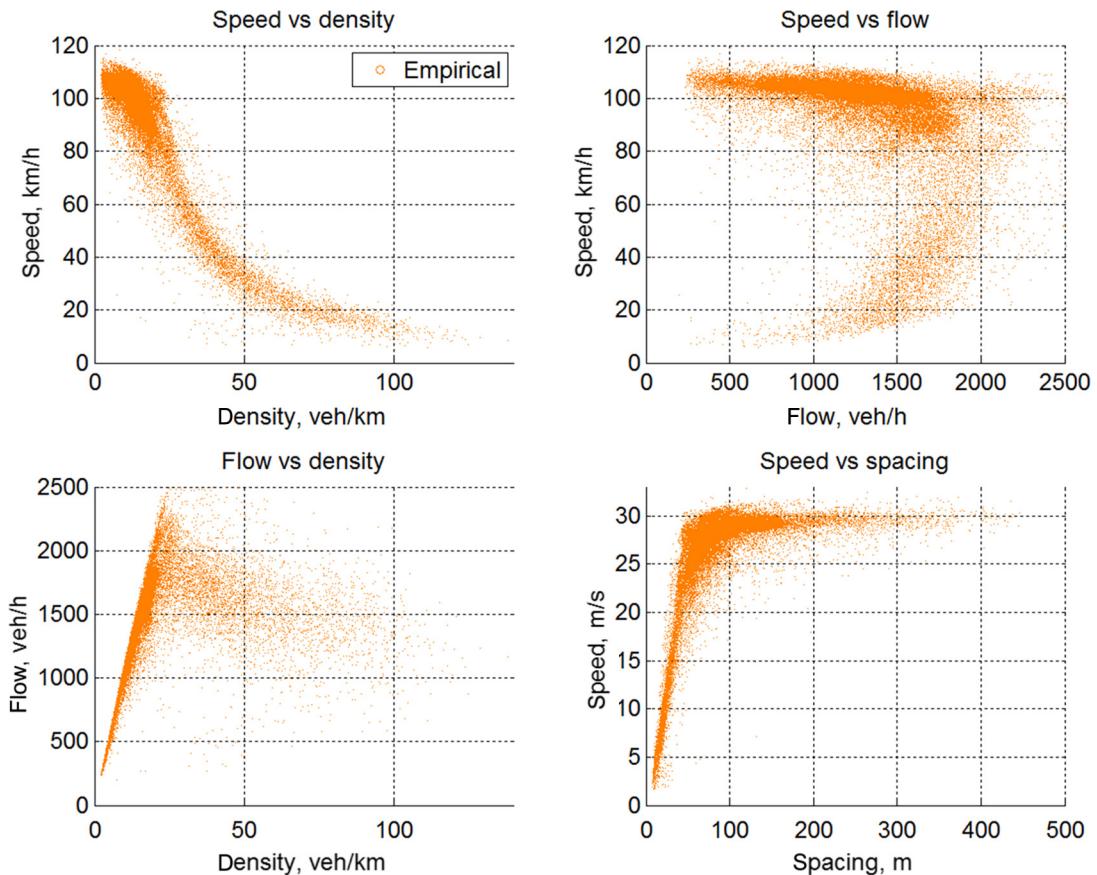


Figure 4.3 Observed q - k - v relationships.

in a stop-and-go condition ($v \rightarrow 0$). Except for the two intercepts, the remaining trend of the speed-density relationship is debatable, having been debated over the years, and is still debated today. Some examples of historical efforts on this subject are given below.

4.1.1 The Greenshields Model

Since the exact relationship between speed and density is unclear, Greenshields [9] proposed the use of a linear function to summarize the speed-density relationship. Such a function can be completely determined from knowledge of two points on the line: ($k = 0, v = v_f$) and ($k = k_j, v = 0$). Hence, the speed-density v - k relationship (illustrated in Figure 4.4) can be expressed as

$$v = v_f \left(1 - \frac{k}{k_j}\right).$$

Combining the identity $q = k \times v$ and eliminating v , one is able to derive the flow-density q - k relationship implied by the Greenshields model (Figure 4.5):

$$q = v_f \left(k - \frac{k^2}{k_j}\right).$$

It is interesting to note a few special points on the curve. When the density is close to zero ($k \rightarrow 0$), the flow drops to zero ($q \rightarrow 0$) since the road is almost empty; when the road is jammed ($k = k_j$), the flow also becomes zero ($q = 0$) because no one can move. In addition, since this is

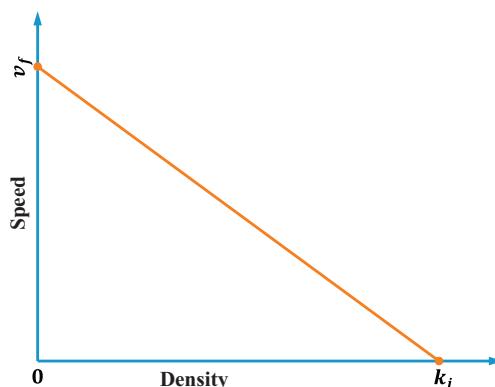


Figure 4.4 The Greenshields speed-density relationship.

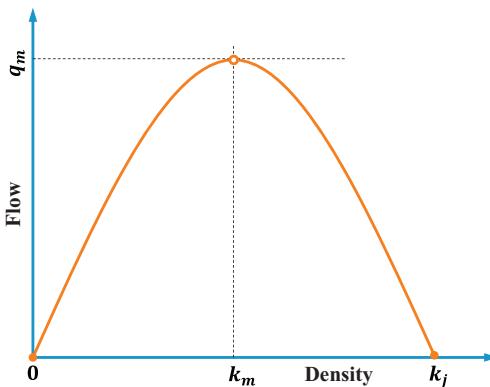


Figure 4.5 Greenshields flow-density relationship.

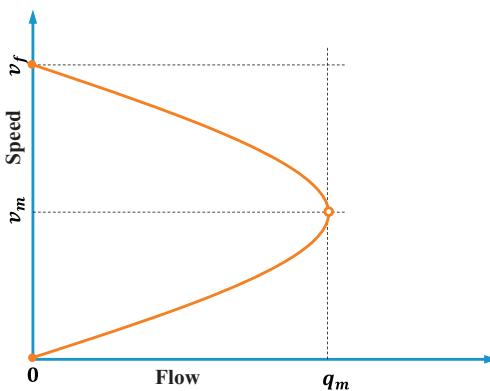


Figure 4.6 Greenshields speed-flow relationship.

a quadratic function with a negative second-order term, the corresponding q - k curve is parabolic with a downward opening. Therefore, starting from the origin ($k = 0, q = 0$), flow increases as density increases. This trend continues until, at some point ($k = k_m$), the flow peaks ($q = q_m = \frac{v_f k_j}{4}$). After this point, the flow begins to drop as the density continues to increase, and the flow becomes zero ($q = 0$) when the density reaches the jam density ($k = k_j$). In this notation, q_m is the maximum flow—that is, the capacity—and k_m is the optimal density—that is, the density when the flow peaks.

Similarly, one can eliminate k from the Greenshields model by using the identity and obtain a speed-flow v - q relationship (Figure 4.6):

$$q = k_j \left(v - \frac{v^2}{v_f} \right).$$

This is again a quadratic function with an opening to the left. When the flow is close zero ($q \rightarrow 0$), two scenarios are possible: (1) the road is nearly empty and the few vehicles on the road are able to move at free-flow speed ($v \rightarrow v_f$); (2) the road is jammed, so that no one is able to move ($v \rightarrow 0$). Actually, entering a given flow value less than the capacity into the equation will normally result in two speeds: a lower one, which corresponds to a worse traffic condition, and a higher one, corresponding to a better traffic condition. When the flow reaches capacity ($q = q_m$), the two speeds become one, which is called the optimal speed, v_m .

Figure 4.7 summarizes the above discussion graphically and puts the speed-density, flow-density, and speed-flow relationships together. Remark-

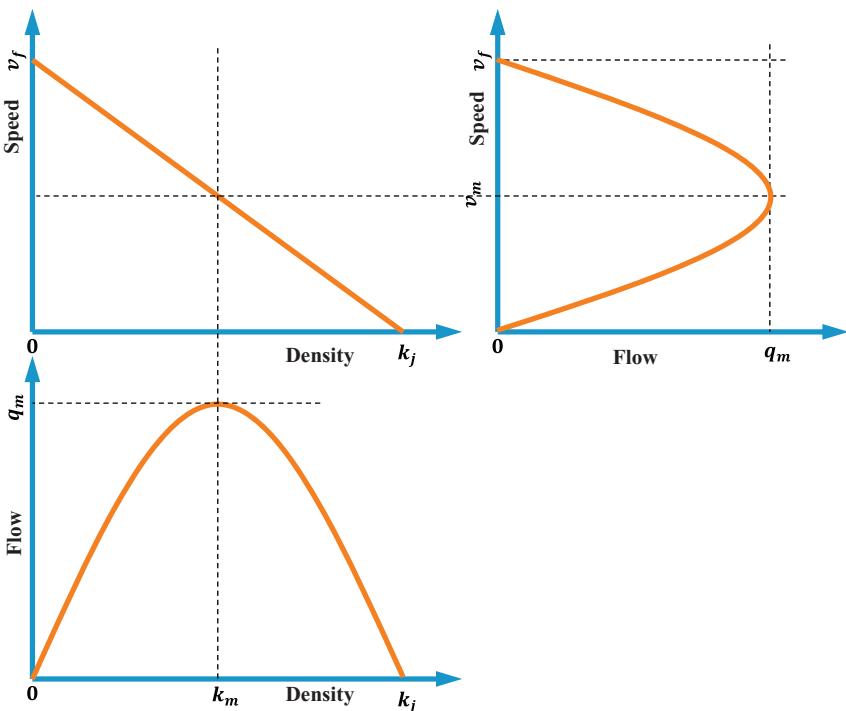


Figure 4.7 Greenshields flow-density-speed relationship.

ably, the relationships among flow, speed, and density such as those depicted in [Figure 4.3](#) and modeled in [Figure 4.7](#) are unique to vehicular traffic flow and are not observed in any other kind of flow, such as gas flow, fluid flow, and flow of Internet packets. Hence, the model and its associated graphical representation that summarizes the pairwise relationships among traffic flow characteristics are the distinguishing features of vehicular traffic flow. Therefore, they are referred to as the *fundamental diagram* in traffic flow theory. The work by Greenshields depicted in [Figure 4.7](#) constitutes the first fundamental diagram in traffic flow theory.

Note that the three pairwise relationships—that is, the speed-density, flow-density, and speed-flow relationships—reflect different facets of the flow-speed-density relationship. Hence, they have different applications in traffic flow theory. For example, the speed-density relationship relates a driver's speed choice to the concentration of vehicles around the driver. Therefore, the relationship is typically used in traffic flow theory to understand how drivers adjust their speeds in response to traffic in their vicinity—that is, modeling drivers' car-following behavior. As will be seen later, the flow-density relationship is convenient for explaining the propagation of disturbances in traffic flow (such as waves and their velocities) and, hence, is frequently used in dynamic traffic flow modeling. Anyone who is familiar with highway capacity and level of service (LOS) will immediately recognize that the speed-flow relationship is extensively used by traffic engineers to perform highway capacity analysis and determine the LOS on freeways and multilane highways.

4.1.2 Other Single-Regime Models

Owing to its simplicity and elegance, the Greenshields model, together with its associated fundamental diagram, is ideal for illustration and pedagogical purposes. Empirical observations reveal that the model suffers from a lack of accuracy, which is salient in [Figure 4.8](#), where the Greenshields model is plotted on top of field observations. For example, the model predicts that the capacity ($q = q_m$) occurs at half the jam density ($k_m = \frac{1}{2}k_j$). If an average vehicle length of 6 m or 20 feet is assumed, the jam density would be somewhere around $1000/6 \approx 164$ vehicles per kilometer or $5280/20 \approx 264$ vehicles per mile. Half of this number is 82 vehicles per kilometer or 132 vehicles per mile. However, field observations suggest that k_m is most likely in the range of 25–40 vehicles per kilometer or 40–65 vehicles per mile. In addition, unlike the way that speed decreases linearly with density, field observations show that free-flow speed can be sustained up to a density

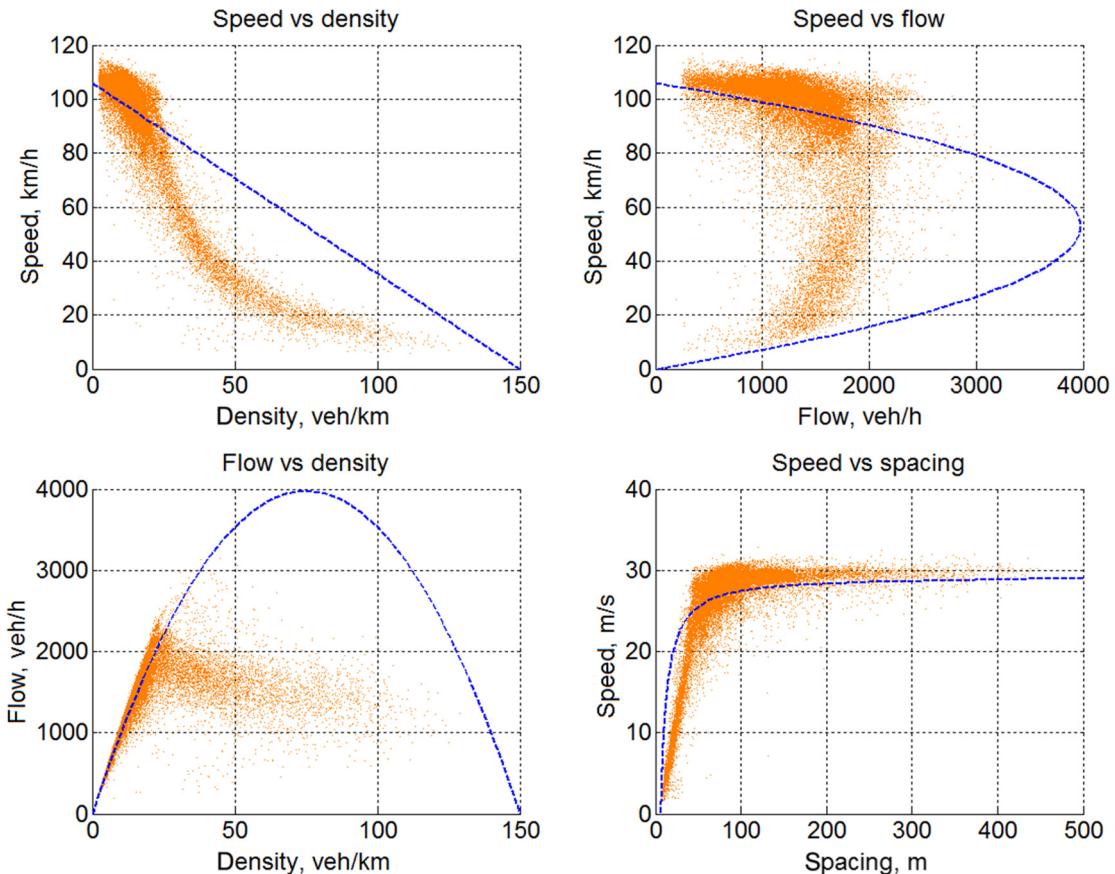


Figure 4.8 Fundamental diagrams implied by Greenshields model.

Table 4.1 Single-regime models

Authors	Model	Parameters
Greenshields [9]	$v = v_f \left(1 - \frac{k}{k_j}\right)$	v_f, k_j
Greenberg [10]	$v = v_m \ln \left(\frac{k_j}{k}\right)$	v_m, k_j
Underwood [11]	$v = v_f e^{-\frac{k}{k_m}}$	v_f, k_m
Drake et al. [12]	$v = v_f e^{-\frac{1}{2} \left(\frac{k}{k_m}\right)^2}$	v_f, k_m
Drew [13]	$v = v_f \left[1 - \left(\frac{k}{k_j}\right)^{n+\frac{1}{2}}\right]$	v_f, k_j, n
Pipes [14] and Munjal [15]	$v = v_f \left[1 - \left(\frac{k}{k_j}\right)^n\right]$	v_f, k_j, n

v_f is free-flow speed, k_j is jam density, v_m is optimal speed, k_m is optimal density, and n is an exponent.

of about 15 vehicles per kilometer or 25 vehicles per mile before a noticeable speed drop can be observed.

Inspired by Greenshields's pioneering work, many models were proposed subsequently to formulate speed-density relationships with various degrees of fitting quality. Table 4.1 provides an incomplete list of these early models.

The models in Table 4.1 share one thing in common—they are one-equation models, meaning that the models apply to the entire range of density. Hence, these models are called *single-regime models*. Figure 4.9 shows the performance of these single-regime models by plotting them on top of empirical observations, just to provide some visual feedback of how they approximate reality.

4.2 MULTIREGIME MODELS

It seems that none of these single-regime models are able to fit the empirical observations reasonably well over the entire density range. Some models are good in one density range, while others are superior in another range. The inability of single-regime models to perform well over the entire range of density prompted researchers to think about fitting the data in a piecewise manner using multiple equations. This gave rise to *multiregime models*, an incomplete list of which is given in Table 4.2. Among the list are the Edie model [16], the two-regime linear model, the modified Greenberg model, and the three-regime model.

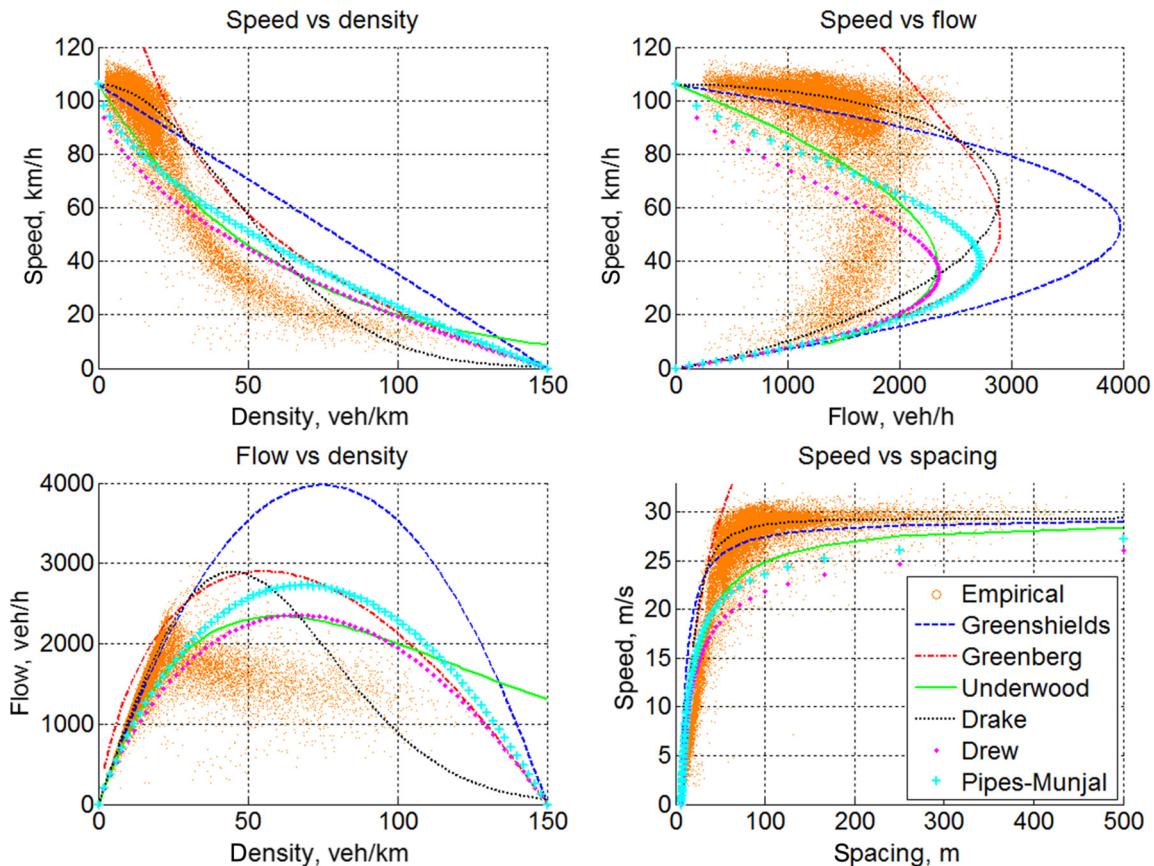


Figure 4.9 Comparison of single-regime models.

Table 4.2 Multi-regime models

Regimes models	Free flow	Transitional	Congested
Edie model	$v = 108e^{-k/163.9}$ $k \leq 20$	- -	$v = 47\ln(162.5/k)$ $k > 20$
Two-regime model	$v = 108 - 0.515k$ $k \leq 30$	- -	$50 - 0.33k$ $k > 30$
Modified Greenberg model	$v = 103$ $k \leq 20$	- -	$v = 52\ln(150/k)$ $k > 20$
Three-regime model	$v = 108 - 0.5k$ $k \leq 20$	$v = 120 - 1.5k$ $20 < k \leq 65$	$v = 40 - 0.256k$ $k > 65$

May [17] presented a comparison of these multiregime models. Similar work illustrated in Figure 4.10 is found in Ref. [18].

4.3 THE STATE-OF-THE-ART MODELS

Early equilibrium models such as the Greenshields model [9], the Greenberg model [10], the Underwood model [11], the Drake model [12], the Drew model [13], and the Pipes-Munjal model [14, 15] are typically simple because they involve only two (the first four models) or three (the last two models) parameters. In addition, they are single-regime models whose derivatives of flow with respect to density ($\frac{dq}{dk}$) exist at each point in the entire range of density. This makes these models mathematically appealing because $\frac{dq}{dk}$ can be very useful later in dynamic macroscopic modeling such as in solving the LWR model (see Chapter 8). Moreover, these macroscopic models are closely related to a family of microscopic car-following models, and we shall revisit such a connection in Section 14. Unfortunately, these models typically suffer from poor fitting quality, as can be seen in Figure 4.9. Multiregime models such as the Edie model [16], the two-regime linear model, the modified Greenberg model, and the three-regime model may come with a little improvement in fitting quality, but their piecewise formulation makes them less attractive.

Further research emphasizes single-regime models, which are apparently coupled with the development of microscopic car-following models. Details of these car-following models and their associated equilibrium models will be discussed in Part III. Highlighted below are a set of more recent equilibrium models.

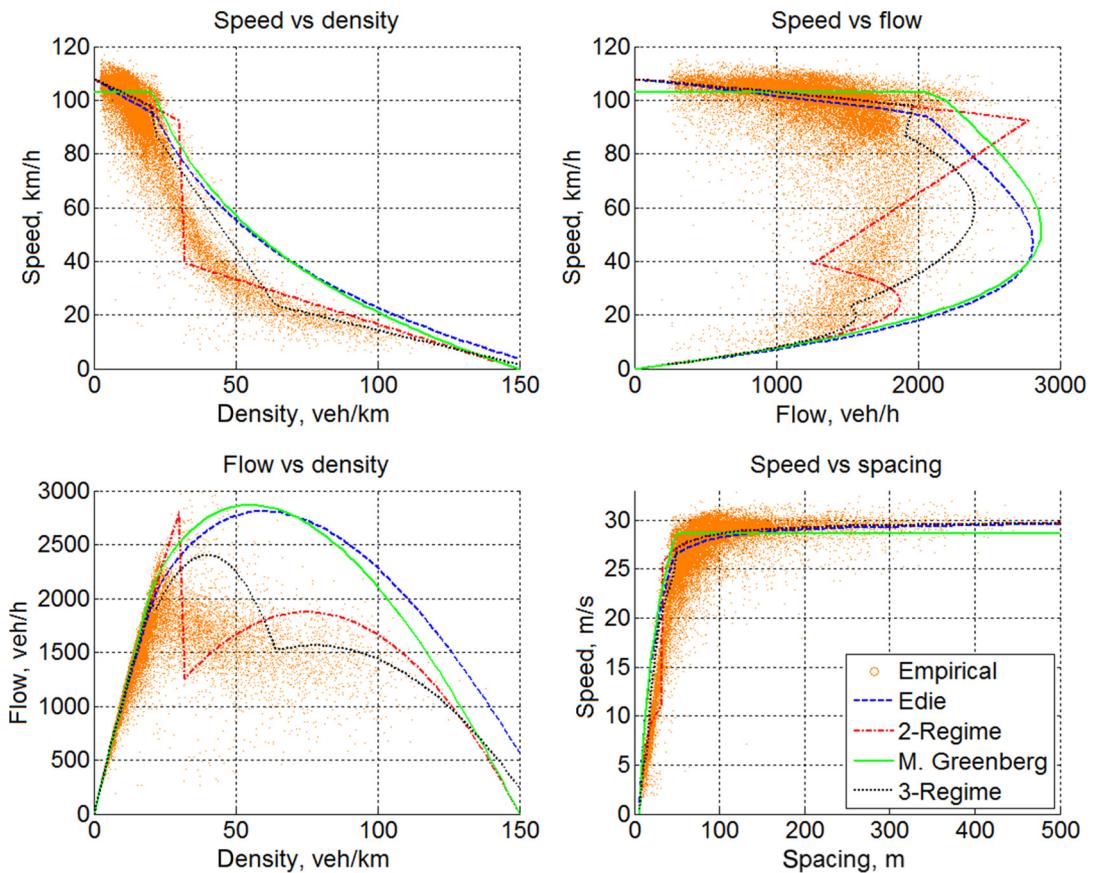


Figure 4.10 Comparison of multiregime models.

Newell Nonlinear Model

The Newell nonlinear model [58] involves three parameters and takes the following form:

$$\nu = \nu_f \left(1 - e^{-\frac{\lambda}{\nu_f} (\frac{1}{k} - \frac{1}{k_j})} \right),$$

where ν_f is the free-flow speed, k_j is the jam density, and λ is the slope of the speed-spacing curve.

Del Castillo and Benítez Model

Also involving three parameters, the model of del Castillo and Benítez [118, 131] takes the following form:

$$\nu = \nu_f \left(1 - e^{1 - e^{\frac{|C_j|}{\nu_f} (\frac{k_j}{k} - 1)}} \right),$$

where ν_f is the free-flow speed, k_j is the jam density, and C_j is the kinematic wave speed at the jam density.

Del Castillo Negative Power Model

Continuing the above effort, del Castillo [120] proposed a new set of models recently, among which the negative power model is reproduced below:

$$\varphi = [(\nu_f \rho)^{-\omega} + (1 - \rho)^{-\omega}]^{-1/\omega},$$

where $\rho = \frac{k}{k_j}$ and $\varphi = \frac{q}{q_0}$, where q_0 is the reference flow, and $\nu_f = -\frac{\nu_f}{C_j}$. As such, this model involves five parameters: reference flow q_0 , jam density k_j , free-flow speed ν_f , kinematic wave speed at jam density C_j , and ω .

Unlike other models in this subsection which result from the corresponding car-following models, the models of del Castillo [118, 120, 131] do not have their microscopic counterparts.

Van Aerde Model

The Van Aerde model [62, 63] involves four parameters and takes the following form:

$$k = \frac{1}{c_1 + c_3 \nu + c_2 / (\nu_f - \nu)},$$

where $c_1 = \frac{\nu_f}{k_j \nu_m^2} (2\nu_m - \nu_f)$, $c_2 = \frac{\nu_f}{k_j \nu_m^2} (\nu_f - \nu_m)^2$, and $c_3 = \frac{1}{q_m} - \frac{\nu_f}{k_j \nu_m^2}$. As such, the parameters of this model are the free-flow speed ν_f , the optimal speed ν_m , the capacity q_m , and the jam density k_j .

Intelligent Driver Model

The intelligent driver model [60, 61] involves four parameters and takes the following form:

$$k = \frac{1}{(s_0 + \nu T)[1 - (\frac{\nu}{\nu_f})^\delta]^{-1/2}}.$$

where parameters are free-flow speed ν_f , jam distance s_0 , safe time headway T , acceleration exponent δ .

Longitudinal Control Model

The longitudinal control model [113] involves four parameters and takes the following form:

$$k = \frac{1}{(\gamma \nu^2 + \tau \nu + l)[1 - \ln(1 - \frac{\nu}{\nu_f})]},$$

where ν_f is the free-flow speed, l is the nominal vehicle length (which is the reciprocal of the jam density, $l = \frac{1}{k_j}$), τ is the perception-reaction time, and γ is the aggressiveness.

To illustrate their features, the above models are fitted to empirical data. Although no effort is made to optimize the parameters, the following general principles apply when one is fitting the models: (1) fix the free-flow speed ν_f of all the models to roughly the same value observed in the data, (2) fix the jam density k_j of all the models to roughly the same value observed in the data, and (3) fix the capacity to roughly the same value observed in the data by tweaking the remaining parameters. The resulting parameter values are listed in [Table 4.3](#) and the fitted models are illustrated in [Figure 4.11](#).

Table 4.3 Model parameters

Models	Parameters
Newell model	$\nu_f = 106 \text{ km/h}; k_j = 167 \text{ vehicles/km}; \lambda = 1.25 \text{ 1/s}$
Del Castillo and Benítez model	$\nu_f = 106 \text{ km/h}; k_j = 167 \text{ vehicles/km}; C_j = 20 \text{ km/h}$
Negative power model	$\nu_f = 106 \text{ km/h}; k_j = 167 \text{ vehicles/km}; C_j = -16.56 \text{ km/h}; \omega = 50$
Van Aerde model	$\nu_f = 106 \text{ km/h}; k_j = 167 \text{ vehicles/km}; \nu_m = 20 \text{ km/h}; q_m = 2400 \text{ veh/h}$
Intelligent driver model	$\nu_f = 106 \text{ km/h}; s_0 = 6 \text{ m}; T = 1.25 \text{ s}; \delta = 15$
Longitudinal control model	$\nu_f = 106 \text{ km/h}; l = 6 \text{ m}; \tau = 1.3 \text{ s}; \gamma = -0.04 \text{ s}^2/\text{m}$

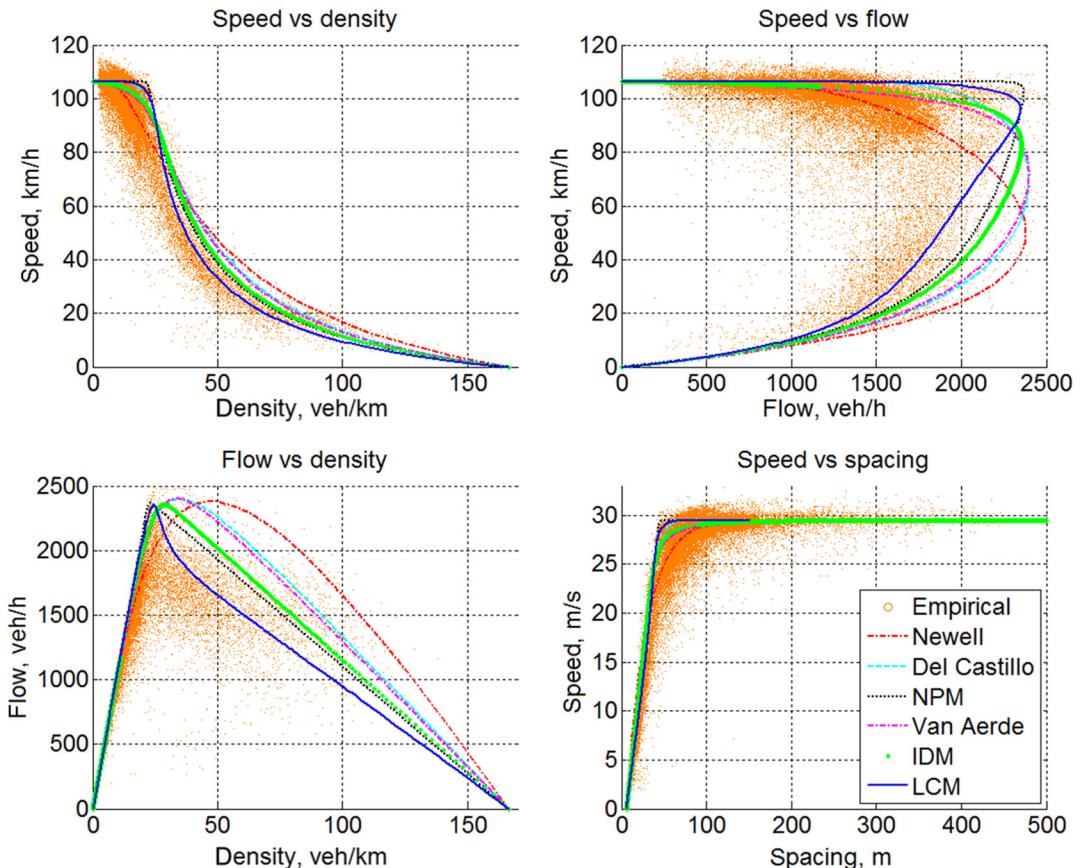


Figure 4.11 State-of-the-art models fitted to empirical data

Although many criteria have been proposed to evaluate fitting quality and a few algorithms have been developed to optimize the fitting process, the following reality check can be used as a visual inspection of the fitting quality:

1. **Free-flow speed** can be directly read off speed-density and speed-flow diagrams as the y -intercept. It is also the slope of the tangent to the flow-density curve that passes through the origin. It is typically not an issue to meet this criterion since one can easily estimate a value from empirical data and fix the model parameter free-flow speed v_f directly at this value.
2. **Jam density** can also be estimated from empirical data by following the trend of the tail of the scatter plot in the speed-density or flow-density diagram. Consequently, this value can be used to set the model parameter jam density k_j or the nominal vehicle length $l = \frac{1}{k_j}$. With the above two ends fixed, it is a good test of fitting quality to examine the capacity condition, which constitutes the third point of interest between the above two points.
3. The **capacity condition** includes the following checkpoints:
 - a. *The location of the capacity* (q_m , k_m , v_m), where capacity q_m is the peak of the flow-density or speed-flow curve, and k_m and v_m are the optimal density and the optimal speed at capacity, respectively.
 - b. *The shape of the curve around the capacity* may exhibit the following types:
 - Skewed parabola: typically observed in outer-lane traffic
 - Triangular: typically observed in middle-lane traffic
 - Reverse-lambda shape: typically observed in inner-lane traffic

In Figure 4.11, the location of the capacity can be easily identified in speed-flow and flow-density diagrams as the tips of the curves. In addition, all three curve shapes can be found. Note that longitudinal control model can even be configured to exhibit the reverse-lambda shape in addition to the other two types.

4.4 CAN WE GO ANY FURTHER?

Though all relationships presented above take deterministic forms, the actual relationships are essentially quite random. For example, a speed-density relationship may predict that when the density k is 12 vehicles per kilometer or 20 vehicles per mile, the speed v will be 96 km/h or 60 miles per hour. However, in reality, the observed speed may vary over a certain range, forming a distribution (see Figure 4.12). The significance of these models lies in their ability to predict a value that makes statistical sense. For example, if one observes traffic for sufficiently long and collects enough speed samples,

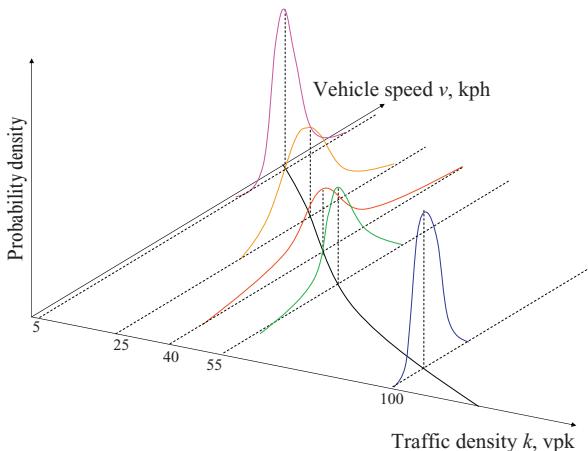


Figure 4.12 Three-dimensional representation of the speed-density relationship.

the likelihood of having a speed in the neighborhood of 96 km/h or 60 miles per hour is very high. Figures 4.9 and 4.10 illustrate the scattering effect of empirical observations and how deterministic models fail to capture such an effect.

Therefore, a step forward to advance the modeling of the speed-density relationship and hence its associated fundamental diagram is to consider the scattering effect by representing speed as a distribution at each density level (see Figure 4.12). Empirical observations seem to support such a proposition. For example, in Figure 4.13 the observed mean and standard deviation of the speed-density relationship are plotted in a single figure. Hence, the deterministic speed-density relationship in the form

$$v = f(k)$$

may be replaced by the following one in generic form:

$$v = f(k, \omega(k)),$$

where ω is a distribution parameter dependent (at least) on density k . In this model, since speed will be a distribution at each density level, the model is essentially a stochastic one. Readers are referred to [18–20] for attempts to obtain stochastic speed-density relationships.

The above pairwise relationships (i.e., equilibrium models) will become handy in the next chapter when we are setting up equations for macroscopic modeling and later for solving the LWR model.

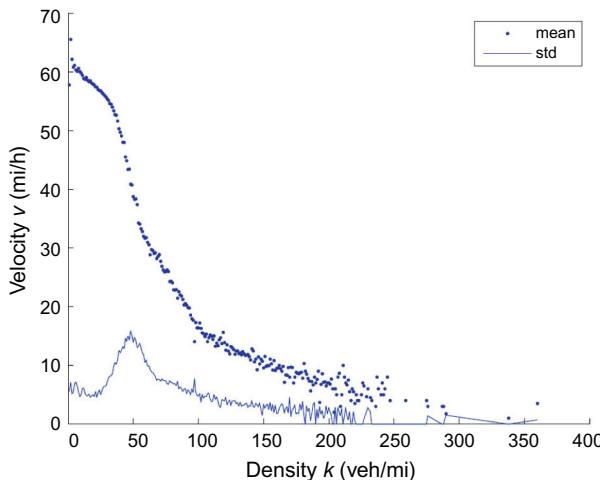
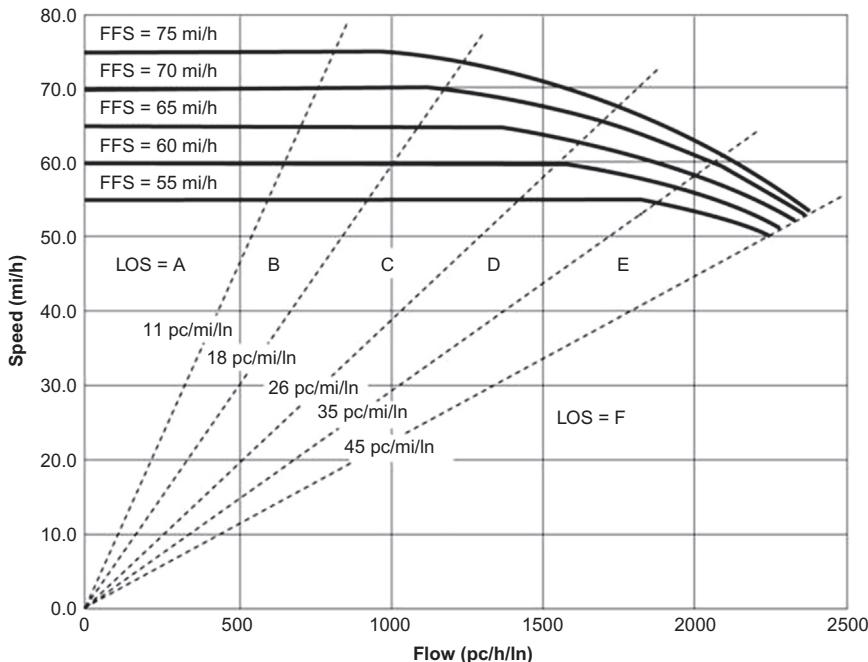


Figure 4.13 Mean and variance of the speed-density relationship.

PROBLEMS

1. From the linear speed-density relationship $v = v_f (1 - k/k_j)$, derive flow-density and speed-flow relationships. With these relationships, find the capacity q_m and the optimal speed v_m and density k_m when the capacity is reached.
2. Study of traffic flow characteristics on a segment of Massachusetts Turnpike (Interstate 90) shows that the following flow-density relationship holds: $q = 65k - 0.36k^2$ pc/h/ln.
 - a. Find the optimal speed, optimal density, and the capacity of this highway segment.
 - b. Comment on how realistic the capacity is.
 - c. In addition, find the speed when the highway is at half of its capacity.
 - d. Comment on whether the above result makes sense.
3. The figure below is the speed-flow relationship used in the HCM to determine the LOS on a basic freeway segment and multilane highways. Use the curve labeled “FFS = 70 mi/h” to do the following:
 - a. Find the free-flow speed v_f indicated by this curve.
 - b. Find the capacity q_m indicated by this curve.
 - c. Find the optimal speed v_m indicated by this curve.
 - d. Estimate the optimal density k_m with use of the identity.
 - e. What LOS does this k_m correspond to?



- f. If the Greenshields model applies, calculate the jam density k_j , and further optimal density k_m .
- g. What LOS does this k_m corresponds to?
- h. Comment on how the Greenshields model approximates the HCM curve.
4. Derive the capacity q_m implied by the Greenberg model and find its associated optimal density k_m .
5. Derive the capacity q_m implied by the Underwood model and find its associated optimal speed v_m .
6. An engineering student estimated a free-flow speed of 60 miles per hour and a capacity of 3600 vehicles per hour on a section of highway. For a given period, a space mean speed of 45 miles per hour was estimated. If the Underwood model applies, what would you estimate the flow rate of this period to be?
7. On a section of Interstate 91 near the University of Massachusetts Amherst, studies show that the speed- density relationship is $v = v_f[1 - ((k/k_j)^{2.5})]$. Assume a capacity of 4600 vehicles per hour and that the

jam density is 200 vehicles per mile. Find the free-flow speed and the optimal speed at capacity.

8. Payne [21] proposed an early empirical speed-density relationship and used this relationship in his macroscopic traffic simulation model FREFLO:

$$\nu = \min\{88.5, (172 - 3.72k + 0.0346k^2 - 0.00119k^3)\},$$

where ν is in kilometers per hour and k is in vehicles per kilometer.

- a. Plot the speed-density relationship graphically (you may draw it manually, do it in Excel, or use a computer program such as MATLAB). Use the plot to do the following:
- b. Identify the free-flow speed ν_f .
- c. Identify the valid range of the density k in this model—that is, the range of k that yields nonnegative speed. Label the jam density k_j as the upper bound of this range.
- d. Identify the capacity condition—that is, capacity q_m , optimal speed ν_m , and optimal density k_m .

CHAPTER 5

Conservation Law

In previous chapters, two types of relationships among traffic flow characteristics were discussed

1. The flow-speed-density relationship or the identity,

$$q = k \times v.$$

Note that (1) it is an identity—that is, it is self-guaranteed by the generalized definition of traffic flow characteristics; (2) it is location specific and time specific— $q(t, x) = k(t, x) \times v(t, x)$ —that is, flow, speed, and density must refer to the same location and time.

2. Pairwise relationships or equilibrium models,

$$v = V(k),$$

$$q = Q(k),$$

$$v = U(q).$$

Note that (1) they define the fundamental diagram and hence differentiate vehicular traffic flow from other kinds of flows; (2) they are location specific—that is, different locations and roads may have different underlying fundamental diagrams; (3) they are equilibrium models—that is, they describe a steady-state behavior in the long run, and hence are not specific to a particular time; (4) such relationships are only of statistical significance—that is, the equal signs do not strictly hold in the real world. On the basis of points (2) and (3), these relationships may also be expressed as follows:

$$v(x) = V(k(x)),$$

$$q(x) = Q(k(x)),$$

$$v(x) = U(q(x)).$$

The main purpose of formulating a traffic flow theory is to help better understand traffic flow and, by the application of such knowledge, to control traffic for safer and more efficient operations. Hence, a good theory should be able to help answer the following questions:

- Given existing traffic conditions on a road and upstream arrivals in the near future, how do road traffic conditions change over time?
- Where are the bottlenecks, if any?
- In the case of congestion, how long does it last and how far do queues spill back?
- If an incident occurs, what is the best strategy for cleanup so that the impact on traffic is minimized?

Answers to these questions involve the analysis of dynamic change of traffic states over time and space. Unfortunately, the above relationships or models are capable only of describing traffic states. They do not provide a mechanism to analyze how such states evolve. Starting from this chapter, dynamic models will be introduced to address these questions.

The derivation of a dynamic equation starts with the examination of a small volume of roadway traffic as a continuum. Here traffic flow is treated as a one-dimensional compressible fluid like a gas. Conservation laws apply to this kind of fluid, and the first-order form of conservation is mass conservation, also known as the continuity equation.

5.1 THE CONTINUITY EQUATION

There are several ways to derive the continuity equation, each takes a different perspective on the small volume of roadway traffic (see [Figure 5.1](#)).

Derivation I: Finite Difference

The following derivation is found in Ref. [3]. Suppose a highway section is delineated by two observation stations at x_1 and x_2 . Let $\Delta x = x_2 - x_1$ denote the section length. During time interval $\Delta t = t_2 - t_1$, N_1 vehicles passed x_1 and N_2 vehicles passed x_2 . Therefore, the flow rates at these locations are

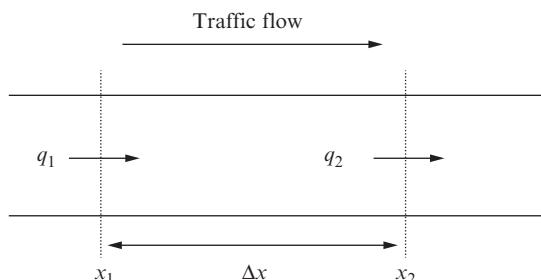


Figure 5.1 Deriving the continuity equation I.

$$q_1 = \frac{N_1}{\Delta t} \quad \text{and} \quad q_2 = \frac{N_2}{\Delta t}.$$

The change in the number of vehicles in the section is

$$\Delta N = N_2 - N_1 = (q_2 - q_1)\Delta t = \Delta q\Delta t.$$

Assume the traffic densities in the section at t_1 and t_2 are k_1 and k_2 , respectively. Therefore, there are $M_1 = k_1\Delta x$ vehicles in the section at time t_1 and $M_2 = k_2\Delta x$ vehicles in the section at time t_2 . Alternatively, the change in the number of vehicles in the section can be expressed as

$$\Delta M = k_1\Delta x - k_2\Delta x = (k_1 - k_2)\Delta x = -\Delta k\Delta x.$$

Since vehicles cannot be created or destroyed inside the section, the change in the number of vehicles should be the same in the same section during the same time interval. Therefore, $\Delta N = \Delta M$ —that is,

$$\Delta q\Delta t = -\Delta k\Delta x,$$

$$\Delta q\Delta t + \Delta k\Delta x = 0.$$

Dividing both sides by $\Delta x\Delta t$, we get

$$\frac{\Delta q}{\Delta x} + \frac{\Delta k}{\Delta t} = 0.$$

If we let $\Delta x \rightarrow 0$ and $\Delta t \rightarrow 0$, the above difference equation becomes a partial differential equation:

$$\frac{\partial q}{\partial x} + \frac{\partial k}{\partial t} = 0.$$

The above equation can be abbreviated as

$$q_x + k_t = 0,$$

where $q_x = \frac{\partial q}{\partial x}$ and $k_t = \frac{\partial k}{\partial t}$.

Derivation II: Finite Difference

The derivation is basically the same as above, but is presented in a slightly different way. [Figure 5.2](#) sketches a highway section $\Delta x = x_2 - x_1$ during time interval $\Delta t = t_2 - t_1$. At time t_1 , there are N_1 vehicles in the section and at time t_2 , there are N_2 vehicles in the section. During the period, traffic keeps flowing into the section at rate q_1 and flowing out at rate q_2 . On the basis of vehicle conservation, the following relationship holds:

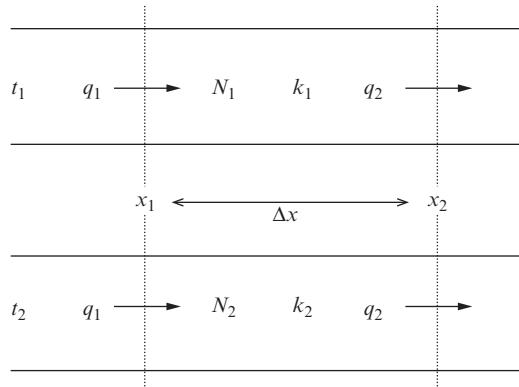


Figure 5.2 Deriving the continuity equation II.

Vehicles at $t_2 = \text{vehicles at } t_1 + \text{inflow during } \Delta t - \text{outflow during } \Delta t$.
This is

$$N_2 = N_1 + q_1 \Delta t - q_2 \Delta t.$$

Note that $N = k\Delta x$, so the above becomes

$$k_2 \Delta x = k_1 \Delta x + q_1 \Delta t - q_2 \Delta t.$$

After arranging terms and dividing both sides by $\Delta x \Delta t$, we get

$$\frac{k_2 - k_1}{\Delta t} = -\frac{q_2 - q_1}{\Delta x}.$$

If we let $\Delta x \rightarrow 0$ and $\Delta t \rightarrow 0$,

$$q_x + k_t = 0$$

Derivation III: Fluid Dynamics

Figure 5.3 illustrates a small fluid cube of size $\delta x \times \delta y \times \delta z$. The fluid velocity v and density k at two sides of the cube also are shown.

The mass flow into the cube is $\nu k \delta y \delta z$. The mass flow out of the cube is as follows:

$$\begin{aligned} (\nu + \delta \nu)(k + \delta k) \delta y \delta z &= \left(\nu + \frac{\partial \nu}{\partial x} \delta x \right) \left(k + \frac{\partial k}{\partial x} \delta x \right) \delta y \delta z \\ &= \left(\nu k + \nu \frac{\partial k}{\partial x} \delta x + k \frac{\partial \nu}{\partial x} \delta x + \frac{\partial \nu}{\partial x} \frac{\partial k}{\partial x} \delta x \delta z \right) \delta y \delta z. \end{aligned}$$

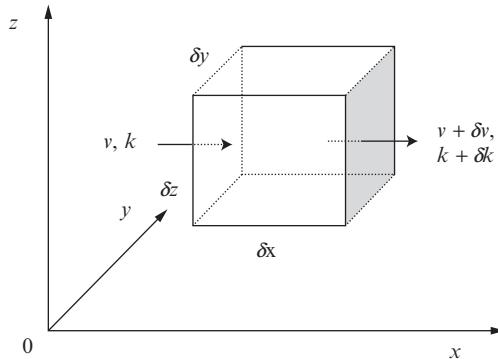


Figure 5.3 Deriving the continuity equation III.

The mass stored in the cube is equivalent to the mass that flows in minus mass that flows out:

$$\left(v \frac{\partial k}{\partial x} \delta x + k \frac{\partial v}{\partial x} \delta x + \frac{\partial v}{\partial x} \frac{\partial k}{\partial x} \delta x \delta z \right) \delta y \delta z = \left(v \frac{\partial k}{\partial x} + k \frac{\partial v}{\partial x} + \frac{\partial v}{\partial x} \frac{\partial k}{\partial x} \delta x \right) \delta x \delta y \delta z.$$

If we ignore the higher-order term, we have

$$\left(v \frac{\partial k}{\partial x} + k \frac{\partial v}{\partial x} \right) \delta x \delta y \delta z = \frac{\partial(kv)}{\partial x} \delta x \delta y \delta z.$$

Similar treatment applies to the other two directions of the cube, so the total mass stored in the cube is

$$\left(\frac{\partial(kv)}{\partial x} + \frac{\partial(ku)}{\partial y} + \frac{\partial(kw)}{\partial z} \right) \delta x \delta y \delta z.$$

The mass stored in the cube must be balanced by the change of mass in the cube:

$$\frac{\partial k}{\partial t} \delta x \delta y \delta z.$$

The law of mass conservation requires that

$$\left(\frac{\partial(kv)}{\partial x} + \frac{\partial(ku)}{\partial y} + \frac{\partial(kw)}{\partial z} \right) \delta x \delta y \delta z + \frac{\partial k}{\partial t} \delta x \delta y \delta z = 0.$$

Therefore,

$$\frac{\partial k}{\partial t} + \left(\frac{\partial(kv)}{\partial x} + \frac{\partial(ku)}{\partial y} + \frac{\partial(kw)}{\partial z} \right) = 0.$$

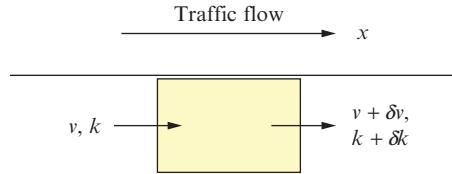


Figure 5.4 Reducing three dimensions to one dimension.

Highway traffic constitutes a special case of the above situation with only one dimension (see Figure 5.4). Using the result derived above, one obtains

$$\frac{\partial(kv)}{\partial x} + \frac{\partial k}{\partial t} = 0.$$

Note that $q = kv$. Therefore,

$$q_x + k_t = 0.$$

Derivation IV: Scalar Conservation Law

This derivation is adopted from [22]. Consider a cell in the time-space domain bounded by $(t_1, t_2) \times (x_1, x_2)$ (see Figure 5.5). Let traffic flow, speed, and density be functions of time and space—that is, $q = q(t, x)$, $v = v(t, x)$, and $k = k(t, x)$. Obviously, the conservation of vehicles in the cell requires the following:

$$\int_{x_1}^{x_2} k(t_2, x) dx - \int_{x_1}^{x_2} k(t_1, x) dx = \int_{t_1}^{t_2} q(t, x_1) dt - \int_{t_1}^{t_2} q(t, x_2) dt,$$

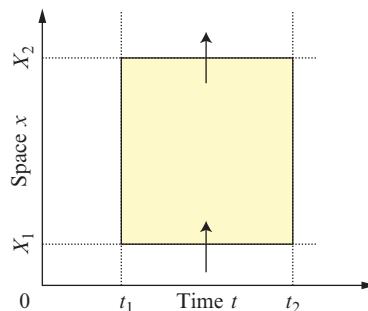


Figure 5.5 Deriving the continuity equation IV.

$$\int_{x_1}^{x_2} [k(t_2, x) - k(t_1, x)] dx = \int_{t_1}^{t_2} [q(t, x_1) - q(t, x_2)] dt.$$

If $k(t, x)$ and $q(t, x)$ are differentiable in x and t , one obtains

$$\begin{aligned} \int_{x_1}^{x_2} \int_{t_1}^{t_2} \frac{\partial k(t, x)}{\partial t} dt dx &= - \int_{t_1}^{t_2} \int_{x_1}^{x_2} \frac{\partial q(t, x)}{\partial x} dx dt, \\ \int_{x_1}^{x_2} \int_{t_1}^{t_2} \left[\frac{\partial k(t, x)}{\partial t} + \frac{\partial q(t, x)}{\partial x} \right] dx dt &= 0. \end{aligned}$$

According to the fundamental theorem of calculus of variables, one obtains

$$\frac{\partial k(t, x)}{\partial t} + \frac{\partial q(t, x)}{\partial x} = 0;$$

that is,

$$q_x + k_t = 0.$$

Derivation V: Three-Dimensional Representation of Traffic Flow

As discussed in Chapter 3, the surface which represents the cumulative number of vehicles, N , can be expressed as a function of time t and space x —that is, $N = N(t, x)$. The density at time-space point (t, x) is the first partial derivative of $N(t, x)$ with respect to x , but takes a negative value:

$$k(t, x) = -\frac{\partial N(t, x)}{\partial x}.$$

The flow at (t, x) is the first partial derivative of $N(t, x)$ with respect to t :

$$q(t, x) = \frac{\partial N(t, x)}{\partial t}.$$

If both the flow and the density have first-order derivatives,

$$\frac{\partial q(t, x)}{\partial x} = \frac{\partial N(t, x)/\partial t}{\partial x} = \frac{\partial N^2(t, x)}{\partial x \partial t}$$

and

$$\frac{\partial k(t, x)}{\partial t} = \frac{-\partial N(t, x)/\partial x}{\partial t} = -\frac{\partial N^2(t, x)}{\partial x \partial t},$$

then

$$\frac{\partial q(t, x)}{\partial x} = -\frac{\partial k(t, x)}{\partial t};$$

that is,

$$q_x + k_t = 0.$$

5.2 FIRST-ORDER DYNAMIC MODEL

Traffic evolution is the process of how traffic states (e.g., flow q , speed v , and density k) evolve over time t and space x given some initial conditions (e.g., $k_0 = k(0, x)$) and boundary conditions (e.g., $q(t) = q(t, x_0)$). One recognizes that time t and space x are independent variables and traffic states are dependent variables—that is, they are functions of time and space ($q = q(t, x)$, $v = v(t, x)$, $k = k(t, x)$). The continuity equations derived above are able to dynamically relate the change of flow q_x to the change of density k_t :

$$q_x + k_t = 0.$$

This equation contains two unknown variables $q(t, x)$ and $k(t, x)$. Since the number of unknown variables is greater than the number of equations, the problem is underspecified. Because of this, another simultaneous equation is needed. Hopefully, the identity comes handy:

$$q(t, x) = k(t, x)v(t, x).$$

By adding a new equation, we introduce a third unknown variable—that is, speed $v(t, x)$. Therefore, a third simultaneous equation is called for. Unfortunately, we are running out of options now since we are unable to find a third governing equation that will definitely hold for any time and space. Consequently, we have to accept the less-than-ideal option by looking at equilibrium traffic flow models (e.g., the Greenshields model), which are known to hold only statistically. Such a model takes the form of

$$v = V(k).$$

Putting everything together, one obtains a system of three equations involving three unknown variables:

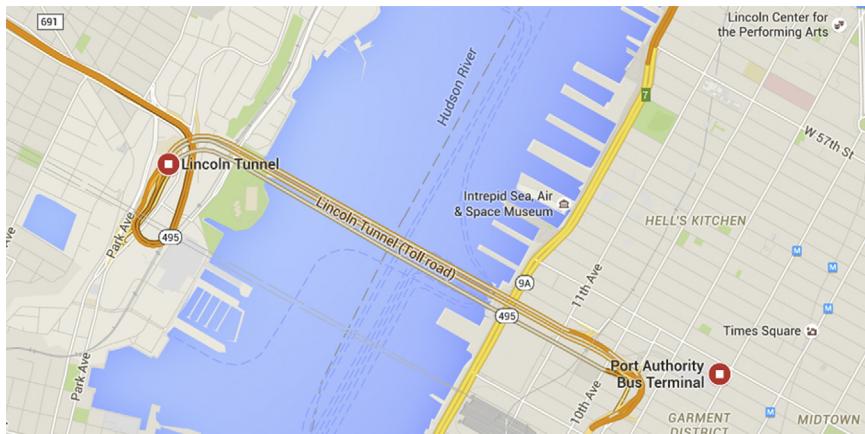
$$\begin{cases} q_x + k_t = 0, \\ q = kv, \\ v = V(k). \end{cases} \quad (5.1)$$

If initial and boundary conditions are provided, the above system of equations may be solvable. If that is the case, one is able to determine the traffic state at an arbitrary time-space point (t, x) —that is, $q(t, x)$, $v(t, x)$, and $k(t, x)$. With such information, one is able to answer the questions posed at the beginning of this chapter.

However, solving such a system of equations is not easy. To make this book self-contained, the following three chapters are designed to help readers ramp up their mathematical knowledge in terms of addressing partial differential equations.

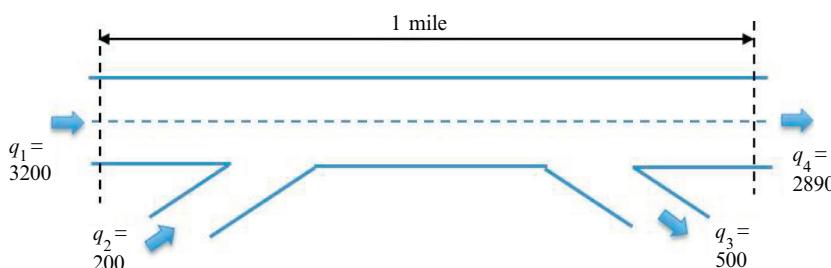
PROBLEMS

- The Lincoln Tunnel is an integral conduit within the New York Metropolitan Area (see the figure below). The tunnel is approximately 1.5 miles (2.4 km) long and consists of three tunnels (north, center, and south) under the Hudson River. A civil engineering consulting firm was contracted to carry out a traffic engineering study on the tunnel, and automatic data collection devices were set up at both ends of the tunnel. The following data were recorded for the south tunnel (number of passenger cars in one direction over two lanes).



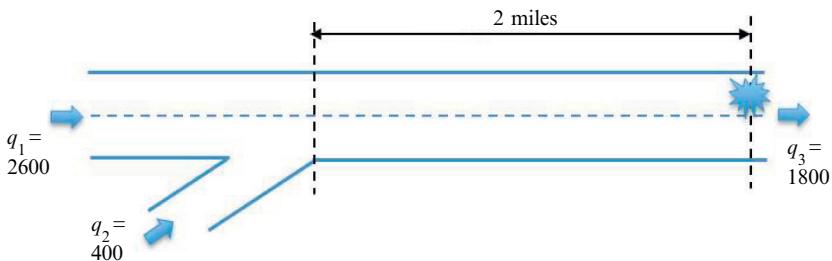
- Find the level of service in each hour with use of the criteria specified by the Transportation Research Board
- A freeway junction includes an on-ramp followed by an off-ramp over a section of 1 mile. At 10:00 a.m., there are 20 vehicles in this section. Assume traffic flows in and out at a rate in vehicles per hour as indicated in the figure below, and calculate the number of vehicles in this section 2 h later.

Time	No. of vehicles that entered	No. of vehicles that exited
00:00	0	0
01:00	90	80
02:00	400	390
03:00	900	874
04:00	1860	1870
05:00	2060	2028
06:00	2200	2210
07:00	3000	2978
08:00	4060	4026
09:00	4200	4154
10:00	3207	3223
11:00	3386	3424
12:00	2810	2832
13:00	3019	3029
14:00	3880	3838
15:00	3665	3637
16:00	4020	3980
17:00	4600	4634
18:00	4282	4316
19:00	3740	3772
20:00	3120	3138
21:00	1680	1706
22:00	408	438
23:00	0	10



3. An accident on Interstate 91 occurred at 8:00 a.m. which blocked one lane, resulting in the remaining lane being capable of discharging traffic at a rate of only 1800 vehicles per hour. Assume that there was no initial queue at the accident location and that traffic keeps arriving from the upstream mainline at a rate of 2600 vehicles per hour and from an on-ramp at a rate of 400 vehicles per hour. The on-ramp is 2 miles from the accident location. Also assume that vehicles maintain a spacing

(i.e., front bumper-to-front bumper distance) of 29.3 feet when they are in a queue. Massachusetts Department of Transportation's goal is to avoid vehicles backing up to the on-ramp. Otherwise, the queue may spill over onto location intersections via the on-ramp, further worsening the situation. Calculate when vehicles will back up to the on-ramp so that the Massachusetts Department of Transportation has a sense of urgency in dispatching a rescue team to clean up the accident.



CHAPTER 6

Waves

To solve the set of equations presented at the end of Chapter 5, one has to leave the topic of traffic flow for a moment and study waves first. One would agree that this is necessary when one looks at [Figure 6.1](#), where vehicle trajectories recorded in the field are plotted on a time-space diagram. The horizontal axis is time, with left being earlier and right later. The vertical axis is space, with traffic flowing upward. Three ripples are clearly visible in this picture depicting the propagation of some disturbances in the traffic. This observation suggests that traffic does behave like waves, and solutions to traffic dynamics can be sought on the basis of the knowledge of waves. As such, the purpose of this chapter is to provide a jump-start introduction to waves.

6.1 WAVE PHENOMENA

Waves are everywhere in the real world. When a pebble is thrown into a pond, one sees ripples circling outward. This is a wave (see [Figure 6.2](#)). When the audience at a football stadium becomes thrilled and rows of the audience stand up and sit down successively, one sees a “signal” bouncing. This is also a wave. When shaking a rope at one end with the other end fixed, one sees a “hump” moving away. This is yet another wave. Basically, *a wave is the propagation of a disturbance in a medium over time and space*. In the above examples, the ripples, signal, and hump are disturbances, while the water, audience, and rope are media. If we apply the notion to a platoon of vehicles on a highway, when one of the vehicles brakes suddenly and then resumes its original speed, subsequent vehicles will be affected successively. The propagation of such a “jerking” effect is a wave, with the jerk being the disturbance and the traffic being the medium. The ripples in [Figure 6.1](#) are examples of such a wave.

6.2 MATHEMATICAL REPRESENTATION

The mathematical language to describe wave phenomena is the partial differential equation (PDE).



Figure 6.1 Traffic waves observed on a highway.



Figure 6.2 Surface waves.

6.2.1 Notation

If a dependent variable k is a function of independent variables t and x , we write $k = k(t, x)$ and we denote its partial derivatives with respect to x and t as follows:

$$k_x = \frac{\partial k}{\partial x}, k_t = \frac{\partial k}{\partial t}, k_{xt} = \frac{\partial^2 k}{\partial x \partial t}, k_{tx} = \frac{\partial^2 k}{\partial t \partial x}, k_{xx} = \frac{\partial^2 k}{\partial x^2}, k_{tt} = \frac{\partial^2 k}{\partial t^2}.$$

A PDE for $k(t, x)$ is an equation that involves one or more partial derivatives of k with respect to t and x . For example,

$$k_t = k_x + k, k_t = k_{xx} + k_x + 5, k_t = k_{xxx} + 4k + \cos x.$$

6.2.2 Terminology

PDEs can be classified on the basis of their order, homogeneity, and linearity.

Order

The order of a PDE is the order of the highest partial derivative in the equation. For example,

- *first-order* PDE: $k_t = k_x + k$;
- *second-order* PDE: $k_t = k_{xx} + k_x + 5$;
- *third-order* PDE: $k_t = k_{xxx} + 4k + \cos x$.

A first-order PDE can be expressed in the following general form:

$$P(t, x, k)k_t + Q(t, x, k)k_x = R(t, x, k),$$

where P , Q , and R are coefficients, and they may be functions of t , x , and k .

Homogeneity

A first-order PDE $P(t, x, k)k_t + Q(t, x, k)k_x = R(t, x, k)$ may be

- *homogeneous* if $R(t, x, k) = 0$;
- *nonhomogeneous* if $R(t, x, k) \neq 0$.

Linearity

In the above general first-order PDE, if both P and Q are independent of k —that is, $P = P(t, x)$, $Q = Q(t, x)$ —and

- If R is also independent of k —that is, $R = R(t, x)$ —then the PDE is *strictly linear*. For example, $2xk_t + 3k_x = 5t$.
- If R is linearly dependent on k —that is, $R = R(t, x, k)$ —then the PDE is *linear*. For example, $2xk_t + 3k_x = 5k + 3$.
- If R is dependent on k in a nonlinear manner, then the PDE is *semilinear*. For example, $2xk_t + 3k_x = e^k$.

In particular, if P or Q is dependent on k , or both P and Q are dependent on k —that is, $P = P(t, x, k)$, $Q = Q(t, x, k)$ —and $R = R(t, x, k)$, then the PDE is *quasilinear*. For example, $k_t + (3k + 2)k_x = 0$.

A PDE is nonlinear if it involves cross terms of k and its derivatives—for example, $k_t k_x + k = 2$.

Now, test yourself by classifying the following PDEs:

1. $k_t + ck_x = 0$.
2. $k_t + ck_x = e^{-t}$.
3. $k_{tt} = C^2 k_{xx}$, where C is a constant.
4. $k_{tt} - k_x x + k = 0$.
5. $k_{tt} + kk_x + k_{xxx} = 0$.

6.3 TRAVELING WAVES

Many PDEs have solutions in a traveling wave form $k(t, x) = f(x - ct)$.¹ Figure 6.3 illustrates two instants of the traveling wave, $f(x - ct_0)$ and $f(x - ct_1)$. It is easy to find that (1) the traveling wave preserves its shape and (2) the wave at time t_1 is simply a horizontal translation of its initial profile at time

¹ The following discussion is derived from Ref. [23] with modifications.

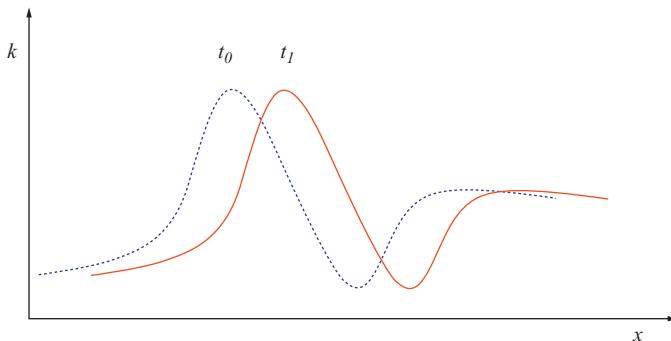


Figure 6.3 A traveling wave.

t_0 . If c is a positive constant, wave $k(t, x) = f(x - ct)$ travels to the right over time, while wave $k(t, x) = f(x + ct)$ moves to the left over time.

6.4 TRAVELING WAVE SOLUTIONS

Solve the following wave equation:

$$k_{tt} = ak_{xx},$$

where a is a constant.

Assume that a solution to the above wave equation takes a traveling form $k(t, x) = f(x - ct)$. Let $z = x - ct$. Then

$$k_t = \frac{\partial k}{\partial t} = \frac{df}{dz} \frac{\partial z}{\partial t} = f' \times (-c) = -cf'.$$

Similarly, $k_x = f'$, $k_{tt} = c^2 f''$, and $k_{xx} = f''$.

Plugging the above expressions into the wave equation, one obtains

$$(c^2 - a)f'' = 0.$$

There are two ways for the left-hand side to be 0: (1) $c^2 - a = 0$ and (2) $f'' = 0$.

1: If $c^2 - a = 0$, then $k(t, x) = f(x \pm \sqrt{a}t)$, where f can take any functional form.

2: If $f'' = 0$, then $k(t, x) = A + B(x - ct)$, where A and B are arbitrary constants.

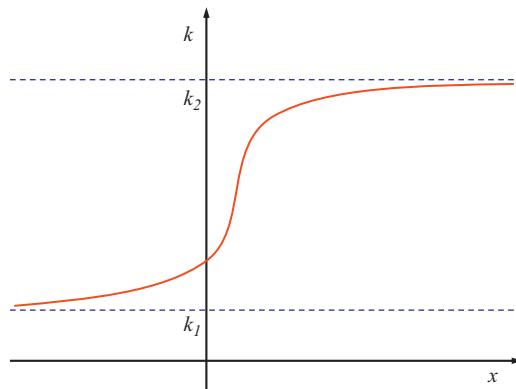


Figure 6.4 Wave front and pulse.

6.5 WAVE FRONT AND PULSE

A traveling wave is called a wave front if

$$\begin{cases} k(t, x) = k_1 & \text{as } x \rightarrow -\infty, \\ k(t, x) = k_2 & \text{as } x \rightarrow +\infty. \end{cases}$$

Figure 6.4 illustrates a wave front. A traveling wave is called a pulse if $k_1 = k_2$.

6.6 GENERAL SOLUTION TO WAVE EQUATIONS

Many wave equations have a general solution in the form of superposition of traveling waves:

$$k(t, x) = F(x - ct) + G(x + ct).$$

Note that even though each of the terms on right-hand side is a traveling wave, their superposition may not necessarily be.

Example 1

Solve the following wave equation with initial conditions

$$\begin{cases} k_{tt} = c^2 k_{xx}, \\ k(x, 0) = f(x), \\ k_t(x, 0) = g(x), \\ -\infty < x < +\infty, t > 0. \end{cases}$$

Solution

Applying the above general solution to the first initial condition, we have

$$F(x) + G(x) = k(x, 0) = f(x).$$

Applying the above general solution to the second initial condition, we have

$$-cF'(x) + cG'(x) = k_t(x, 0) = g(x).$$

Dividing both sides by c and integrating, we obtain

$$-F(x) + G(x) = -F(0) + G(0) + \frac{1}{c} \int_0^x g(s)ds.$$

Solving for $F(x)$ and $G(x)$, we obtain

$$\begin{cases} F(x) = \frac{1}{2}f(x) - \frac{1}{2}[-F(0) + G(0) + \frac{1}{c} \int_0^x g(s)ds], \\ G(x) = \frac{1}{2}f(x) + \frac{1}{2}[-F(0) + G(0) + \frac{1}{c} \int_0^x g(s)ds]. \end{cases}$$

Plugging the result back into the general solution, we obtain

$$\begin{aligned} k(t, x) &= F(x - ct) + G(x + ct) \\ &= \frac{1}{2}f(x - ct) - \frac{1}{2} \left[-F(0) + G(0) + \frac{1}{c} \int_0^{x-ct} g(s)ds \right] \\ &\quad + \frac{1}{2}f(x + ct) + \frac{1}{2} \left[-F(0) + G(0) + \frac{1}{c} \int_0^{x+ct} g(s)ds \right]. \end{aligned}$$

We combine terms and we obtain a specific generic solution:

$$k(t, x) = \frac{1}{2}[f(x - ct) + f(x + ct)] + \frac{1}{2c} \int_{x-ct}^{x+ct} g(s)ds.$$

This is called the *d'Alembert solution*.

Example 2

Solve the following wave equation with initial conditions

$$\begin{cases} k_{tt} = 4k_{xx}, \\ k(x, 0) = e^{-x^2}, \\ k_t(x, 0) = 0, \\ -\infty < x < +\infty, t > 0. \end{cases}$$

Solution

Applying the result in [Example 6.6](#) and considering that $k_t(x, 0) = g(x) = 0$, one obtains

$$k(t, x) = \frac{1}{2}[f(x - ct) + f(x + ct)].$$

Therefore,

$$k(x, 0) = \frac{1}{2}[f(x) + f(x)] = f(x) = e^{-x^2}$$

Since $f(x) = k(x, 0)$, one obtains

$$k(t, x) = \frac{1}{2}[k(x - ct, 0) + k(x + ct, 0)] = \frac{1}{2}[e^{-(x-ct)^2} + e^{-(x+ct)^2}].$$

6.7 CHARACTERISTICS

Consider [Example 6.6](#), since $f(x) = k(x, 0)$ and $g(x) = k_t(x, 0)$, the solution can be transformed to the following form:

$$k(t, x) = \frac{1}{2}[k(x - ct, 0) + k(x + ct, 0)] + \frac{1}{2c} \int_{x-ct}^{x+ct} g(s)ds.$$

6.7.1 Domain of Dependence

Applying the above conclusion, one notices that the solution k at an arbitrary time-space point (t^*, x^*) is

$$k(t^*, x^*) = \frac{1}{2}[k(x^* - ct^*, 0) + k(x^* + ct^*, 0)] + \frac{1}{2c} \int_{x^*-ct^*}^{x^*+ct^*} g(s)ds.$$

The above equation suggests that the solution at an arbitrary point (t^*, x^*) can be determined by the initial condition at points $(0, x^* - ct^*)$ and $(0, x^* + ct^*)$ and the interval **I** bounded by the two points (inclusive)—that is, $I = [x^* - ct^*, x^* + ct^*]$. This is illustrated in the left part of [Figure 6.5](#). Therefore, the interval **I** is called the *domain of dependence* of point (t^*, x^*) .

6.7.2 Range of Influence

The term “*range of influence*” refers to a collection of time-space points whose solutions are influenced either completely or partially by the domain of dependence **I**; see the shaded area in the right part of [Figure 6.5](#).

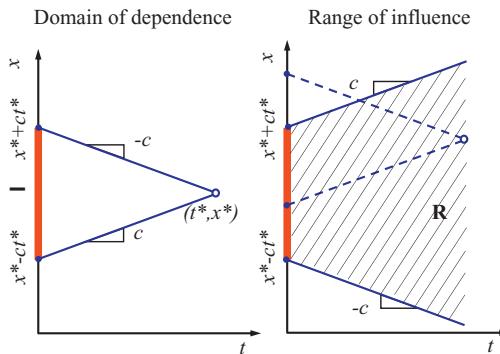


Figure 6.5 Characteristics.

6.7.3 Characteristics

Notice that in the left part of Figure 6.5, the two lines coming from point (t^*, x^*) intersecting the x -axis at $(x^* - ct^*, 0)$ and $(x^* + ct^*, 0)$ have slopes c and $-c$. These two lines are called *characteristic lines* or simply *characteristics* (please do not mix this up with traffic flow characteristics).

6.8 SOLUTION TO THE WAVE EQUATION

In a special case where $k_t(0, x) = 0$, the solution of the wave equation in Example 6.6 reduces to

$$k(t, x) = \frac{1}{2}[k(0, x - ct) + k(0, x + ct)].$$

This shows that the value of k at (t, x) depends only on the initial values of k at two points, $x_1 = x - ct$ and $x_2 = x + ct$. Once the initial values $k(0, x - ct)$ and $k(0, x + ct)$ are known, one constructs the solution k at (t, x) by taking the average of $k(0, x_1)$ and $k(0, x_2)$.

Example 3

Use characteristics to solve the following wave equation:

$$\begin{cases} k_{tt} = 4k_{xx}, \\ k(0, x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \text{ or} \\ 0 & \text{otherwise,} \end{cases} \\ k_t(0, x) = 0, \\ -\infty < x < +\infty, t > 0. \end{cases}$$

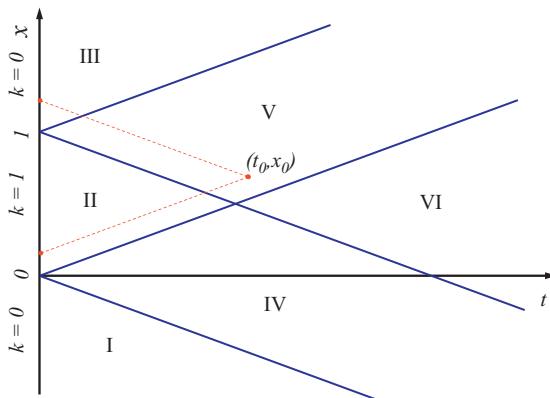


Figure 6.6 Solution to Example 6.8.

In this equation, the traveling wave speed $c = \pm 2$ —that is, $k(t, x) = f(x \pm 2t)$. First, one constructs an x - t plane. Locate points 0 and 1 on the x -axis. Then one draws two characteristics (their slopes are ± 2) from each of the two points (see Figure 6.6). The four characteristics partition the x - t plane into six regions as labeled in Figure 6.6. Take an arbitrary point (t_0, x_0) , for example. The solution at this point is found by drawing two characteristics from this point. Then find the intersections of the two characteristics on the x -axis. Next, find the k values at the two intersections. In this case the k values are 1 and 0. Then the solution k at point (t_0, x_0) is the average of the k values at the two intersections—that is, $k(t_0, x_0) = \frac{1}{2}$.

With use of a similar technique, the solution in other regions can be determined. To sum up, the solution to the above wave equation is as follows:

$$k(t, x) = \begin{cases} 0 & \text{if } (t, x) \in \text{region I}, \\ 1 & \text{if } (t, x) \in \text{region II}, \\ 0 & \text{if } (t, x) \in \text{region III}, \\ \frac{1}{2} & \text{if } (t, x) \in \text{region IV}, \\ \frac{1}{2} & \text{if } (t, x) \in \text{region V}, \\ 0 & \text{if } (t, x) \in \text{region VI}. \end{cases}$$

The above discussion presents the following notion:

1. For some wave equations such as that in Example 6.8, solution k at point (t, x) can somehow be related to the initial condition k_0 at point $(0, x_0)$.

2. This is done by drawing two lines, called characteristics, from (t, x) with slopes c and $-c$.
3. These characteristics intersect the x -axis at two points $(0, x_1)$ and $(0, x_2)$, where $x_1 = x - ct$ and $x_2 = x + ct$. Then the solution is $k(t, x) = \frac{1}{2}[k(0, x_1) + k(0, x_2)]$.

6.9 METHOD OF CHARACTERISTICS

Now let us consider a very simple PDE derived from the conservation law with an initial condition. In Chapter 5, the conservation law led to the following continuity equation:

$$k_t + q_x = 0.$$

If one assumes $q = ck$, where c is a constant, then $q_x = ck_x$, and the PDE can be defined as follows (please ignore the physical meaning of k and q for the moment—this issue will be revisited later):

$$\begin{cases} k_t + ck_x = 0, \\ k(0, x) = k_0(x), \\ -\infty < x < \infty, 0 < t, \\ c \text{ is a constant.} \end{cases}$$

The goal is to find a solution to this PDE, or equivalently find the value of k at an arbitrary time-space point, $k(t, x)$. Rather than working on an arbitrary point in the entire time-space plane, one starts with a simpler case by working on a point on a specific curve in the time-space plane. To do this, one draws a curve $x = x(t)$ (how to draw this curve will be made clear shortly), and the new goal is to find the value of k at an arbitrary point $(t, x(t))$ on the curve—that is, $k(t, x(t))$. To find the solution, let us examine how k changes along the curve $x = x(t)$. The rate of change of k with time is the first (and total) derivative of k with respect to time t ; that is,

$$\frac{dk(t, x(t))}{dt} = \frac{\partial k}{\partial t} \frac{dt}{dt} + \frac{\partial k}{\partial x} \frac{dx(t)}{dt} = k_t + \frac{dx}{dt} k_x.$$

If one compares the right-hand side of this equation with the left-hand side of the original PDE, one recognizes that they are very similar. Actually, they will be identical if one imposes

$$\frac{dx(t)}{dt} = c.$$

Consequently, one obtains

$$\frac{dk(t, x(t))}{dt} = k_t + ck_x = 0.$$

This means that the total time derivative of k along the curve $x = x(t)$ is zero—that is, the value of k is constant along the curve. This implies that the curve $x = x(t)$ needs to be drawn such that it is a straight line with slope of c . Therefore, one finds the equation of the line by solving the following ordinary differential equation:

$$\begin{cases} \frac{dx(t)}{dt} = c, \\ x(0) = x_0. \end{cases}$$

This yields

$$\begin{cases} x(t) = ct + x_0, \\ x_0 = x - ct. \end{cases}$$

At time $t = 0$, this line intersects the x -axis at x_0 . Since k remains constant along this line, the solution k at any point on this line, $k(t, x(t))$, is the same as $k(0, x_0) = k_0(x_0)$, which is given in the initial condition. Therefore, we have found the solution for all points on this line. Such a line is called a *characteristic*. Sounds familiar? Yes, it has the same meaning as the characteristic in previous sections, where it is a line drawn from a time-space point with slope c , which is the speed of the traveling wave $f(x - ct)$.

With the above knowledge, it is simple to find the solution at an arbitrary point, $k(t^*, x^*(t^*))$. The procedure is as follows:

1. Construct the equation of the characteristic drawn from this point: $x(t) = ct + x_0$.
2. Find the intercept of this characteristic on the x -axis: $x_0 = x^* - ct^*$.
3. Find the value of k at the intercept from the initial condition: $k(0, x_0) = k_0(x_0) = k_0(x^* - ct^*)$.
4. Apply this value of k to the point of interest: $k(t^*, x^*) = k_0(x^* - ct^*)$.

Example 4

Use the method of characteristics to find the solution to the following PDE at point $(t^* = 3, x^* = 10)$:

$$\begin{cases} k_t + 2k_x = 0, \\ k(0, x) = 2x^2 + 5, \\ -\infty < x < \infty, 0 < t. \end{cases}$$

Solution

Following the above procedure, one obtains the following:

1. The characteristic drawn from this point is $x(t) = 2t + x_0$.
 2. The intercept of this characteristic on the x -axis is $x_0 = 10 - 2 \times 3 = 4$.
 3. The value of k at the intercept is $k(0, 4) = 2 \times 4^2 + 5 = 37$.
 4. Therefore, $k(3, 10) = 37$.
-

6.10 SOME PROPERTIES

The above discussion is based on a very simple first-order, linear, homogeneous PDE. It is informative to examine further the method of characteristics and note some of its properties.

6.10.1 Properties of Characteristics

In the above example, the characteristic is a straight line, and this is so because c is a constant. Similarly, another characteristic drawn from another time-space point is also a straight line. In addition, the two straight lines are parallel since they have the same slope c . [Figure 6.7](#) illustrates a family of characteristics (in the x - t plane) which are straight and parallel. Each characteristic carries a constant k value denoted by a line above which is labeled as the characteristic curve. Different characteristics may carry different k values, so the surface $k(t, x)$ is not necessarily flat. A *kinematic wave* is a family of characteristics which carry and propagate signals, such as those characteristics illustrated in [Figure 6.7](#).

Now, what if c is not a constant? The following are two examples.

Example 5

In this example, c depends on k but not explicitly on t and x —that is, $c = c(k(t, x))$. In this case, the characteristic equation needs to be derived from

$$\frac{dx(t)}{dt} = c(k(t, x)).$$

Hence, the characteristic equation is

$$x = c(k_0(x_0))t + x_0.$$

Therefore, the characteristic is still a straight line. However, the slope of the line may take different values at different intercepts x_0 . Consequently, two characteristics may intersect. See [Figure 6.8](#) for an illustration.

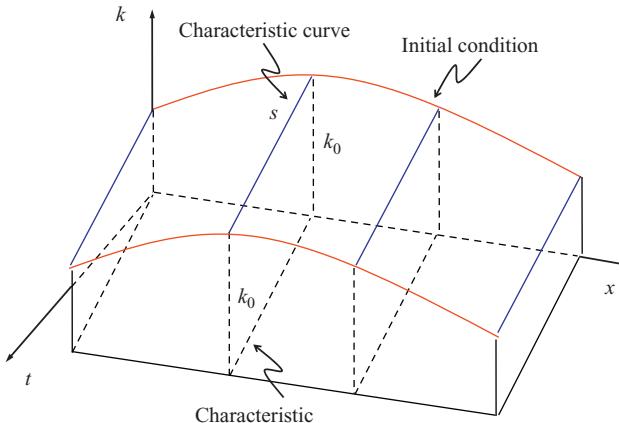


Figure 6.7 Illustration of parallel characteristics.

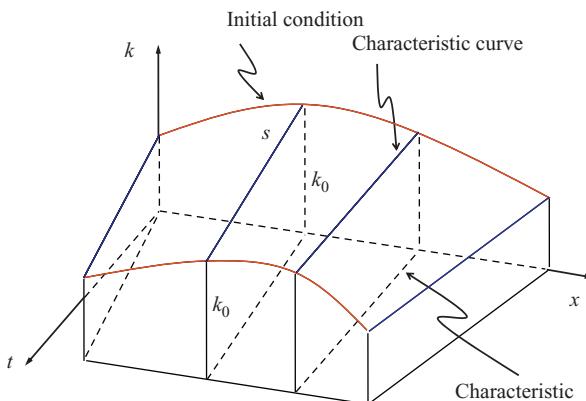


Figure 6.8 Illustration of nonparallel characteristics.

Example 6

In this example, c explicitly depends on x and or t —for example, $c = t$. The equation of the characteristic is derived from the following ordinary differential equation:

$$\frac{dx(t)}{dt} = c = t.$$

After integration, one obtains $x = \frac{1}{2}t^2 + A$, where A is an integral constant. In this case, the characteristic is no longer a straight line, but is a parabola. In addition, characteristics drawn from different time-space points are no longer parallel. Instead, they may intersect.

Summing up the above discussion on characteristics, we have the following:

- If c is a constant, characteristics are straight, parallel lines.
- If c depends on k but not explicitly on t and x , characteristics are still straight lines, but different characteristics may have different slopes and hence these characteristics may intersect.
- If c explicitly depends on x and or t , characteristics are neither straight nor parallel. Consequently, these characteristics may intersect.
- Since a characteristic denotes a set of time-space points on which the solution of k remains constant, k may be multivalued at the intersection of two characteristics. Such an occurrence is called a *gradient catastrophe*.

6.10.2 Properties of the Solution

If one imposes $\frac{dx(t)}{dt} = c$, one obtains

$$\frac{dk}{dt} = 0.$$

This implies that the solution of k remains constant on a characteristic $x = x(t)$. This conclusion holds *only* if the underlying PDE is homogeneous—that is,

$$k_t + ck_x = 0.$$

What if the PDE is not homogeneous? For example,

$$k_t + ck_x = -1.$$

In this case, the total derivative of k with respect to t becomes

$$\frac{dk}{dt} = -1.$$

This implies that k is no longer constant along characteristic $x = x(t)$, but rather linearly decreases at the rate of 1—that is, $k = k_0 - t$, where k_0 is found in the initial conditions. [Figure 6.9](#) illustrate such a case.

PROBLEMS

1. Classify the following partial differential equations:
 - $k_{txt} = 3xk_{tt} + 4t k k_x k_t - 8xt$.
 - $k_t = 9k_x$.
 - $k_{xx} + \frac{1}{5}k_x + \frac{1}{25}k_{tt} = 0$.
 - $kk_x + ak_t + bk = 0$, where a and b are constants.
 - $5k_t + 9k_x = 3k^3$.

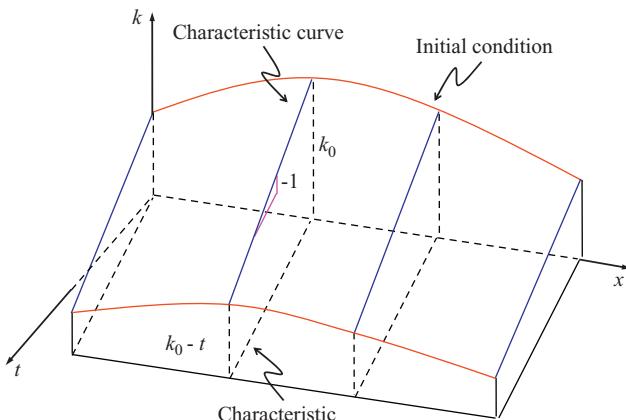


Figure 6.9 Solution of a nonhomogeneous PDE.

2. Use characteristics to find a solution to the following PDE with initial conditions:

$$k_{tt} - k_{xx} = 0,$$

$$\text{where } k(0, x) = \begin{cases} 2 & \text{when } x > 1, \\ 1 & \text{when } -1 \leq x \leq 1 \text{ and } k_t(0, x) = 0, \\ 0 & \text{when } x < -1. \end{cases}$$

3. Find the d'Alembert solution to the following PDE with initial conditions

$$k_{tt} - \frac{1}{9}k_{xx} = 0,$$

$$\text{where } k(0, x) = 0 \text{ and } k_t(0, x) = 2.$$

4. Find the d'Alembert solution to the following PDE with initial conditions

$$k_{tt} = 4k_{xx},$$

$$\text{where } k(0, x) = 2x \text{ and } k_t(0, x) = e^{-x}.$$

5. Use the method of characteristics to solve the following first-order homogeneous linear PDE with an initial condition at time-space point $(t, x) = (4, 5)$:

$$\begin{cases} k_t + \frac{1}{2}k_x = 0, \\ k(0, x) = 4x + \ln x^2, \\ -\infty < x < \infty, \\ t > 0. \end{cases}$$

6. Use the method of characteristics to solve the following first-order homogeneous quasi-linear PDE with an initial condition at time-space point $(t, x) = (2, 20)$:

$$\begin{cases} k_t + (2k + 1)k_x = 0, \\ k(0, x) = x + 10, \\ -\infty < x < \infty. \\ t > 0. \end{cases}$$

7. Use the method of characteristics to solve the following first-order nonhomogeneous quasi-linear PDE with an initial condition at time-space point $(t, x) = (5, 9)$:

$$\begin{cases} k_t + 2tk_x = 2, \\ k(0, x) = 2x + 1, \\ -\infty < x < \infty, \\ t > 0. \end{cases}$$

CHAPTER 7

Shock and Rarefaction Waves

In the previous chapter, the method of characteristics was discussed as a means to solve the continuity equation (i.e., conservation law) with an initial condition:

$$\begin{cases} k_t + q_x = 0, \\ k(0, x) = k_0(x), \\ -\infty < x < \infty, 0 < t, \end{cases}$$

where $q = Q(k)$ is a function of k . To be consistent with the notation in the previous chapter, the following connection needs to be made:

$$q_x = \frac{\partial q}{\partial x} = \frac{\partial Q(k)}{\partial x} = \frac{dQ}{dk} \frac{\partial k}{\partial x} = Q'(k)k_x = ck_x$$

To find the solution of k at an arbitrary time-space point (t^*, x^*) , $k(t^*, x^*)$, one simply constructs a characteristic $x = ct + x_0$ which starts from (t^*, x^*) and extends back to the x -axis at intercept $(0, x^* - ct^*)$. Since $k((0, x^* - ct^*)) = k_0(x^* - ct^*)$ is given in the initial condition and k remains constant along the characteristic, the solution is

$$k(t^*, x^*) = k_0(x^* - ct^*).$$

7.1 GRADIENT CATASTROPHES

In the above discussion, if c is a constant, characteristics drawn from two different time-space points are straight, parallel lines. Hence, any time-space point lies on one and only one characteristic, and the solution at this point is single valued. However, if $c = c(k)$ is a function of k and not explicitly dependent on x or t , two different characteristics drawn from two time-space points are still straight lines but they may not necessarily be parallel, in which case they may intersect and the solution at this intersection may be multivalued. For example, Figure 7.1 illustrates two such characteristics A_0A_4 and B_0B_4 . As the two characteristics become closer and closer, the gradient (i.e., slope) of the solution profile (represented by the red curves 0, 1, 2, 3, and 4 above the two characteristics) becomes increasingly steep.

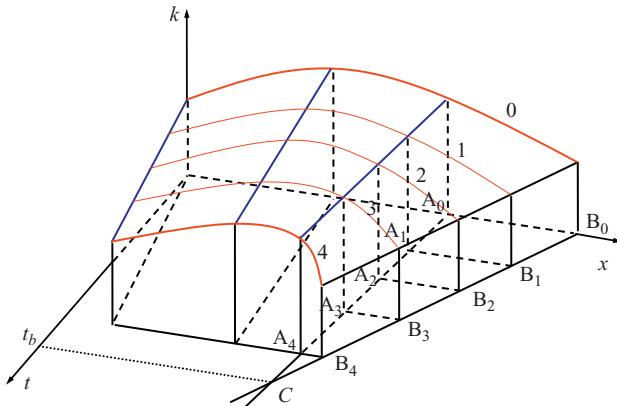


Figure 7.1 A gradient catastrophe.

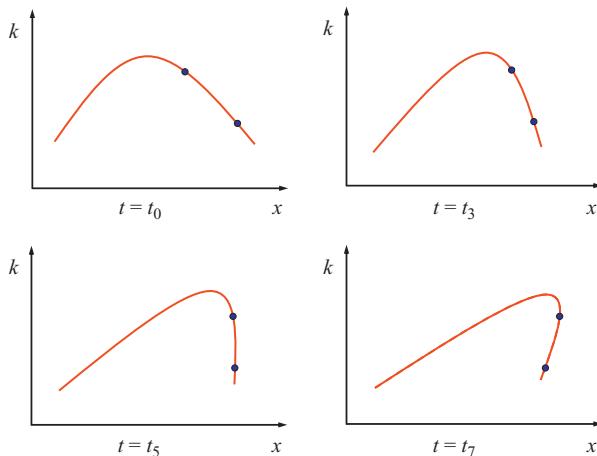


Figure 7.2 Top of profile overtakes bottom of profile.

When the two characteristics intersect at point C , the solution profile will have an infinite gradient at this point. The formation of such an infinite gradient is called a *gradient catastrophe*, and the time when infinite gradient occurs is called the *break time* t_b . Figure 7.2 presents a few frames of time development of the solution profile. Notice that the top dot of the profile moves faster than the bottom dot. Sooner or later, the top dot will catch up with the bottom dot at the break time, creating a gradient catastrophe. After this, the top dot runs over the bottom dot, and the profile ceases to

be a valid function. Consequently, the solution beyond the break time will be problematic. The purpose of this chapter is to address such an issue.

The above example illustrates a family of characteristics moving closer and closer over time, so they form a *compression wave*. The opposite case is a family of characteristics moving farther and farther apart without any intersection (see Figure 7.3); such a wave is called an *expansion wave*. The corresponding time development of the solution profile is shown in Figure 7.4. It can be seen that the bottom dot moves faster than the top dot in this case, and the solution profile becomes thinned out or rarefied.

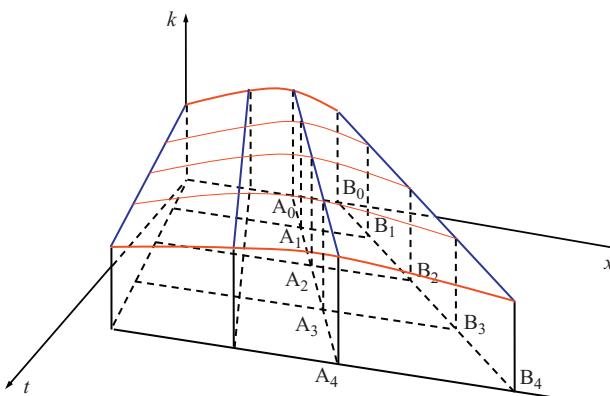


Figure 7.3 Characteristics farther and farther apart.

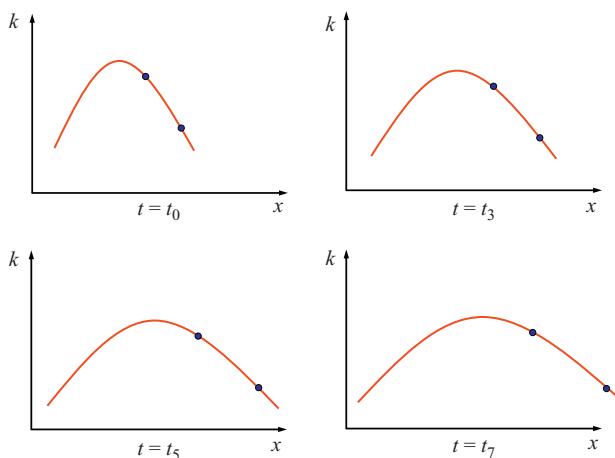


Figure 7.4 Solution profile thinned out.

7.2 SHOCK WAVES

As a continuation of the above discussion, if two characteristics intersect, the solution at the intersection will be multivalued. However, if one allows discontinuity at the intersection, it is possible to construct a piecewise smooth solution. For example, Figure 7.5 illustrates such a solution where curve $x_s(t)$ in the $x-t$ plane is a collection of characteristic intersections. The solution remains constant along each characteristic and terminates at their intersection. Therefore, the curve partitions the solution space into two parts R^- and R^+ and, consequently, separates the solution into two smooth pieces S^- and S^+ . The drop or discontinuity of k at the curve denotes an abrupt change of k which creates a *shock wave*. Such a piecewise smooth solution of the partial differential equation (PDE) is called a shock wave solution.

A critical step in the shock wave solution is to find the curve $x_s(t)$ which connects the intersections of characteristics. Since the curve represents the locations at which a shock wave forms, such a curve is called a *shock path*. In Figure 7.6, two families of characteristics are illustrated where a characteristic may have multiple intersections. Hence, many curves can be drawn by connecting different sets of intersections and, hence, the shock wave may take different paths. Fortunately, the underlying conservation law ensures that only one shock path is valid, and such a shock path must satisfy a physical condition called the *Rankine-Hugoniot jump condition*:

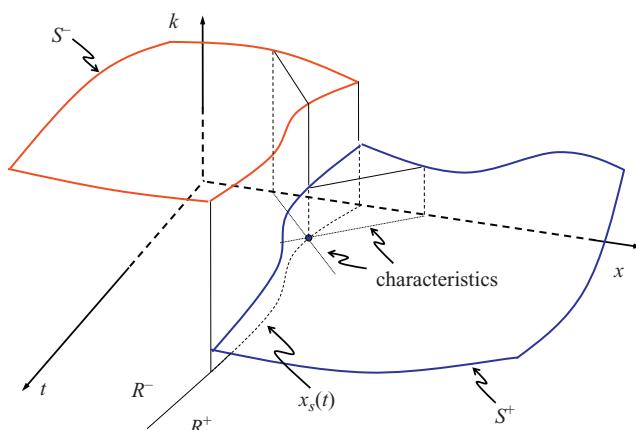


Figure 7.5 Piecewise solution—shock wave.

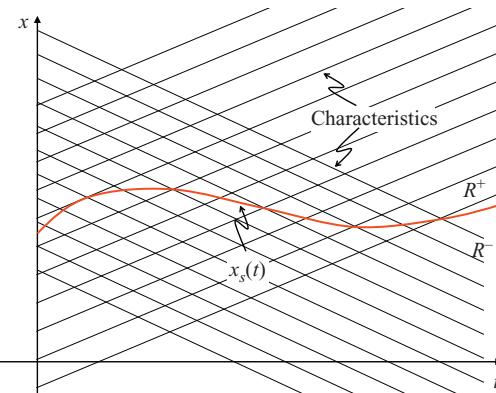


Figure 7.6 Shock path.

$$\frac{dx_s}{dt} = \frac{q(t, x_s^+) - q(t, x_s^-)}{k(t, x_s^+) - k(t, x_s^-)},$$

where $\frac{dx_s}{dt}$ is the slope of the shock path, $q = Q(k)$ as defined in the conservation law, $k(t, x_s^-)$ takes the k value on the R^- side, $k(t, x_s^+)$ takes the k value on the R^+ side, and similar notation applies to $q(t, x_s^-)$ and $q(t, x_s^+)$.

Therefore, if one or more intersections on curve $x_s(t)$ are known, one can construct the shock path by starting from the known points and following the slope defined above.

As an example, solve the conservation law with the following initial conditions:

$$\begin{cases} k_t + q_x = 0, \\ q = \frac{1}{2}k^2, \\ k(0, x) = \begin{cases} 1 & \text{if } x \leq 0, \\ 0 & \text{if } x > 0, \end{cases} \\ -\infty < x < \infty, \\ t > 0. \end{cases}$$

The slope of the characteristics is $c = \frac{dq}{dk} = k$. Obviously, characteristics drawn below $x = 0$ are straight, parallel lines with slope $c = 1$. These characteristics carry the same constant value of $k = 1$, and hence $q = 1/2k^2 = 1/2$. Similarly, characteristics drawn above $x = 0$ are horizontal lines with slope $c = 0$. They carry $k = 0$, and hence $q = 0$. The origin

is a known point on the shock path. According to the Rankine-Hugoniot jump condition, the slope of the shock path is

$$\frac{dx_s}{dt} = \frac{q(t, x_s^+) - q(t, x_s^-)}{k(t, x_s^+) - k(t, x_s^-)} = \frac{0 - 1/2}{0 - 1} = \frac{1}{2}.$$

Therefore, the shock path is a straight line which starts from the origin with constant slope $\frac{1}{2}$ —that is,

$$x_s(t) = \frac{1}{2}t.$$

Therefore, the solution is

$$k(t, x) = \begin{cases} 1 & \text{if } x \leq \frac{1}{2}t, \\ 0 & \text{if } x > \frac{1}{2}t. \end{cases}$$

The solution is illustrated in [Figure 7.7](#). Also illustrated are a few concepts discussed before: a characteristic is a line along which the solution k remains constant; a kinematic wave is a family of straight, parallel characteristics, and a shock wave separates two kinematic waves with an abrupt change of the k value; a shock path is the projection of shock locations onto the $x-t$ plane.

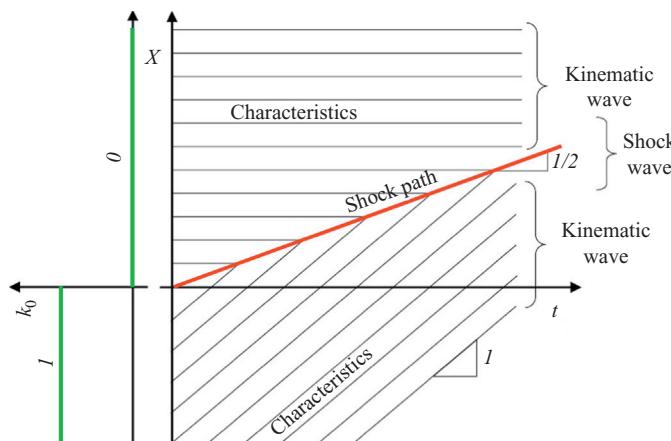


Figure 7.7 An example of a shock path.

7.3 RAREFACTION WAVES

If the initial condition in the above example is reversed¹—that is,

$$k(0, x) = k_0 = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0, \end{cases}$$

characteristics of this PDE should be drawn as in [Figure 7.8](#). In this case, the two families of characteristics go farther and farther apart, leaving an empty wedge-shaped area in between. Since a characteristic carries a constant k solution, areas swept by characteristics will have solutions. An empty area in the solution space means there is no solution in this area. To resolve this issue, there should be a means to fill the empty area with characteristics.

If one relaxes the step function of the initial condition by assuming that k_0 varies smoothly from 0 to 1 over a small distance Δx (see [Figure 7.9](#)), the slopes of characteristics drawn in Δx will gradually increase from 0 to 1 so that any point in the solution space is swept by one and only one characteristic.

To return to the step function of the initial condition, one takes the limit $\Delta x \rightarrow 0$, so [Figure 7.9](#) reduces to [Figure 7.10](#). Now the empty area is filled with a fan of characteristics drawn from the origin. If one cuts the solution space with a few planes $t = t_0, t_1, t_2, \dots$, with t_0 passing the origin and other planes at consequently later times, one obtains a time development of the solution as shown in [Figure 7.11](#). Notice that the profile of the solution is thinned out or rarefied as time moves on. Hence, this fan of characteristics represents a rarefaction wave.

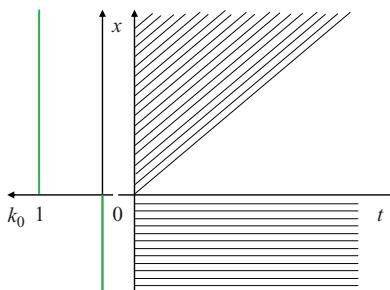


Figure 7.8 Characteristics without an intersection.

¹ The following discussion is derived from Ref. [23] with modifications.

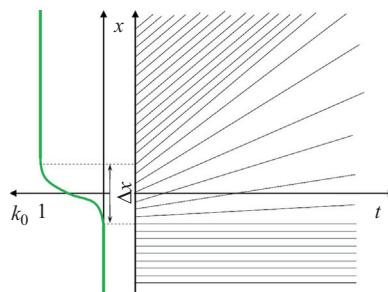


Figure 7.9 Filling an empty area with characteristics.

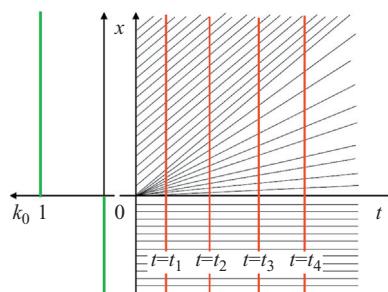


Figure 7.10 A rarefaction wave.

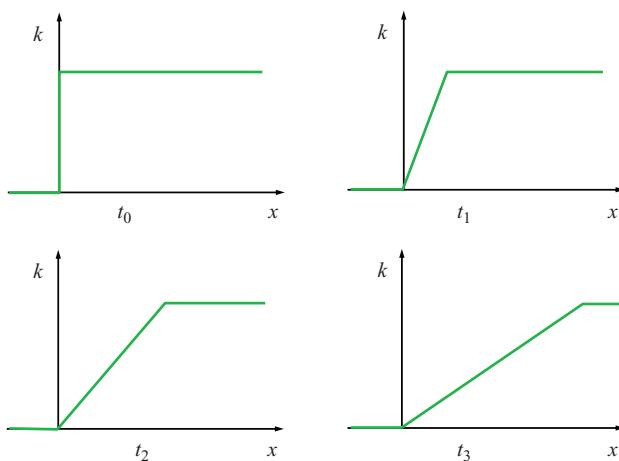


Figure 7.11 Time development of the rarefaction wave.

The rarefaction wave can be used to construct a solution for the following conservation law problem:

$$\begin{cases} k_t + q_x = 0, \\ q = \frac{1}{2}k^2, \\ k(0, x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0, \end{cases} \\ -\infty < x < \infty, \\ t > 0. \end{cases}$$

From the initial condition and with use of the method of characteristics, solutions for the two areas swept by the two parallel characteristics in [Figure 7.8](#) can be easily determined:

$$k(t, x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{if } t < x. \end{cases}$$

The fan of characteristics in the wedge-shaped area consists of lines $x = ct$, where $0 < c < 1$. Therefore, the solution $k(x, t)$ in this area should have the form $f(x/t)$. Hence,

$$k_t = \left(-\frac{x}{t^2}\right)f', \quad k_x = \frac{1}{t}f'.$$

The conservation law can be rewritten as

$$k_t + kk_x = 0.$$

Plugging f and its partial derivatives back into the above equation $k_t + q_x = 0$, we obtain

$$\frac{1}{t}f' \left(f - \frac{x}{t}\right) = 0.$$

If we solve this equation, we get $f'(\frac{x}{t}) = 0$ or $f = \frac{x}{t}$. If $f'(\frac{x}{t}) = 0$, $f(\frac{x}{t}) = k(t, x) = a$, where a is an integral constant. A simple check along $x = 0$ and $x = t$ reveals that this solution does not satisfy the Rankine-Hugoniot jump condition. For example, if we apply the Rankine-Hugoniot jump condition, a shock wave solution bordering area $x \leq 0$ should have a shock path with slope $\frac{a}{2}$. As such, the shock path is a straight line drawn from the origin with a slope of $\frac{a}{2}$. However, none of the characteristics in

area $x \leq 0$ touches the shock path, which makes the shock path an invalid one by definition—that is, a shock wave is the intersection of two or more kinematic waves. Similar reasoning applies to the other side of the wedge bordering area $x > t$. Therefore, solution $f(\frac{x}{t}) = k(t, x) = a$ is not a shock wave solution. The smooth solution is given by $k(t, x) = f(\frac{x}{t}) = \frac{x}{t}$. Hence, the rarefaction wave solution of the original problem is

$$k(t, x) = \begin{cases} 0 & \text{if } x \leq 0, \\ \frac{x}{t} & \text{if } 0 < x \leq t, \\ 1 & \text{if } t < x. \end{cases}$$

One can also construct a shock wave solution by choosing a constant a such that $0 < a < 1$ and applying the Rankine-Hugoniot jump condition:

$$k(t, x) = \begin{cases} 0 & \text{if } x \leq \frac{1}{2}at, \\ a & \text{if } \frac{1}{2}at < x \leq \frac{1}{2}(a+1)t, \\ 1 & \text{if } \frac{1}{2}(a+1)t < x. \end{cases}$$

[Figure 7.12](#) illustrates the solution space which has two shock waves $x = \frac{1}{2}at$ and $x = \frac{1}{2}(a+1)t$. Since the choice of constant a is arbitrary as long as $0 < a < 1$ is met, the above shock wave solution is multivalued. Combining the above rarefaction wave and shock wave solutions, one concludes that the solution to the original problem is not unique. However, the physical process has only one outcome, and hence the solution must be unique. The question is “which solution makes the most physical sense?”

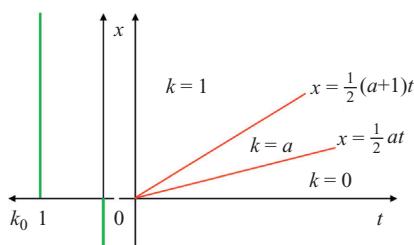


Figure 7.12 A shock wave solution.

7.4 ENTROPY CONDITION

In fluid dynamics, the *entropy condition* is used to select a solution that makes the most physical sense. The entropy condition of a function $k(x, t)$ requires the existence of a positive constant E such that the following inequality is met:

$$\frac{k(t, x + \Delta x) - k(t, x)}{\Delta x} \leq \frac{E}{t}$$

for $\Delta x > 0$ and $t > 0$. Such a condition is shown in Figure 7.13.

Now let us check if a shock wave solution satisfies this condition. We do by slicing the solution space in Figure 7.12 using a plane AA' and projecting the result onto the k - x plane, as shown in Figure 7.14. The solution profile consists of three discontinuous sections with $k = 0, a, 1$. If one choose two arbitrary points Δx apart on the profile as indicated, the slope is $\frac{a}{\Delta x}$. The slope becomes larger and larger as Δx shrinks and becomes infinity at the jump location. Therefore, one cannot find a positive constant E to satisfy the entropy condition, and hence shock wave solutions do not make physical sense.

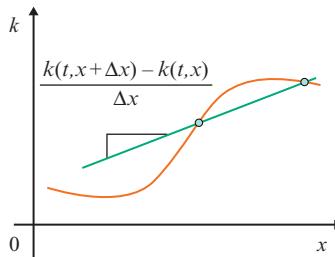


Figure 7.13 Entropy condition.

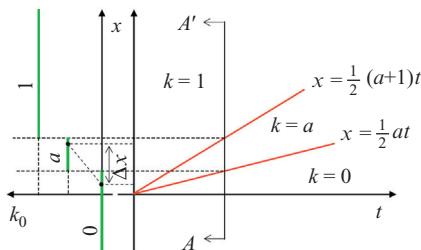


Figure 7.14 Entropy condition in a shock wave solution.

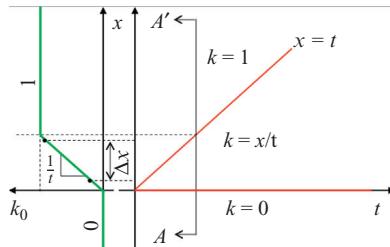


Figure 7.15 Entropy condition in the rarefaction solution.

To apply the same technique to check the rarefaction solution, one slices the solution space with a plane at AA' ; the resultant solution profile is illustrated in Figure 7.15. It can be seen that the maximum slope between any two points on the solution profile is $\frac{1}{t}$. Hence, if one chooses $E = 1$, the entropy condition is always met. Therefore, the rarefaction solution is chosen as the (unique) solution that makes the most physical sense:

$$k(t, x) = \begin{cases} 0 & \text{if } x \leq 0, \\ \frac{x}{t} & \text{if } 0 < x \leq t, \\ 1 & \text{if } t < x. \end{cases}$$

7.5 SUMMARY OF WAVE TERMINOLOGY

At this point, it is helpful to summarize the definition of a few terms frequently used in the analysis of waves and their solutions:

- A *wave* is the propagation in time t and space x of a disturbance in a medium.
- A *signal* is a physical measure (e.g., traffic density k) that describes the disturbance.
- A *characteristic* is a line in the x - t plane along which the signal remains constant.
- A *kinematic wave* is a family of parallel characteristics in the x - t plane.
- A *compression wave* is a family of characteristics which are closer and closer to each other over time.
- A *shock wave* is the formation of an abrupt change in signal in the medium. A compression wave consists of intersecting characteristics. The intersection of these characteristics causes gradient catastrophe, which, in turn, is a precursor of a shock wave.

- An *expansion wave* is a family of characteristics which are farther and farther apart over time.
- A *rarefaction wave* is the effect that the signal profile thins out over time. An expansion wave consists of diverging characteristics which cause two neighboring signals to move farther and farther apart, which, in turn, causes a rarefaction wave.

PROBLEMS

1. Explain the following concepts, and use examples assisted by sketches if necessary.
 - a. Wave
 - b. Characteristic
 - c. Kinematic wave
 - d. Shock wave
2. Find the shock wave solution to the following PDE with initial conditions:

$$\begin{cases} k_t + q_x = 0, \\ q = 2k^2, \\ k(0, x) = \begin{cases} 5 & \text{if } x \leq 0, \\ 2 & \text{if } x > 0, \end{cases} \\ -\infty < x < \infty, \\ t > 0. \end{cases}$$

3. Find the shock wave solution to the following PDE with initial conditions:

$$\begin{cases} k_t + q_x = 0, \\ q = k^3, \\ k(0, x) = \begin{cases} 1 & \text{if } x \leq 2, \\ 0 & \text{if } x > 2, \end{cases} \\ -\infty < x < \infty, \\ t > 0. \end{cases}$$

4. Find the rarefaction wave solution to the following PDE with initial conditions:

$$\begin{cases} k_t + q_x = 0, \\ q = k^2, \\ k(0, x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0, \end{cases} \\ -\infty < x < \infty, \\ t > 0. \end{cases}$$

5. Using the entropy condition to check whether your solution to the above problem is a physically meaningful one.

CHAPTER 8

LWR Model

In previous chapters, we temporarily left traffic flow and concentrated on the conservation law (Chapter 5), waves (Chapter 6), solutions to the conservation law (Chapter 6), and shock waves (Chapter 7). The purpose of these chapters was to pave the road to addressing traffic dynamics and unveiling traffic evolution on highways.

8.1 THE LWR MODEL

At the end of Chapter 5, a dynamic traffic flow model was formulated on the basis of the conservation law:

$$\begin{cases} k_t + q_x = 0, \\ q = kv, \\ v = V(k), \end{cases} \quad (8.1)$$

where $q = q(t, x)$ is flow, $k = k(t, x)$ is density, and $v = v(t, x)$ is mean traffic speed. If one combines the second and third equations by eliminating v , one obtains a flow-density relationship $q = Q(k)$, and the dynamic model becomes:

$$\begin{cases} k_t + q_x = 0, \\ q = Q(k), \end{cases} \quad (8.2)$$

or further

$$k_t + Q'(k)k_x = 0,$$

where $Q'(k) = \frac{dQ(k)}{dk}$. This is the so-called LWR model [24, 25] to honor the three pioneers, Lighthill, Whitham, and Richards, who originally studied this problem. The LWR model is essentially a first-order, homogeneous, quasi-linear partial differential equation.

If we apply the results in the previous chapters, the LWR model with initial condition $k(0, x) = k_0(x)$ can be solved as follows:

1. Construct a time-space diagram (i.e., the t - x plane) with initial condition $k_0(x)$ labeled on the x -axis.

2. Start with an arbitrary point on the x -axis $(0, x^*)$, and determine the k value at this point $k_0(x^*)$ and the value of $c(x^*) = Q'(k_0(x^*))$.
3. Draw a straight line s from point $(0, x^*)$ with slope $c(x^*)$. The line equation is $x_s = c(x^*)t + x^*$, which represents a characteristic along which the k value is constant $k(t, x_s) = k_0(x^*)$.
4. Apply the previous two steps to other points on the x -axis and construct their corresponding characteristics.
5. If two characteristics intersect, terminate both characteristics at their intersection and note the intersection as a point on a shock path. If a characteristic has multiple intersections, use the Rankine-Hugoniot jump condition to determine the right intersection. Repeat this step and find adjacent intersections. Connect these intersections to form a shock path. The solution at both sides of the shock path should be piecewise smooth with a jump along the shock path which forms a shock wave.
6. If two families of characteristics diverge and, hence, leave a wedge-shaped area in between, fill this area with a fan of characteristics and construct a rarefaction wave solution in this area.
7. If an area has multiple rarefaction solutions, apply the entropy condition to select a solution that makes the most physical sense.
8. After the above steps have been followed, the solution space should be filled with characteristics. Each point in the solution space should be swept by one and only one characteristic.
9. If an arbitrary point (t, x) is of interest, one simply follows its characteristic all the way back to the x -axis and reads $k_0(x)$ off the initial condition. This $k_0(x)$ is the k value at the time-space point in question. Consequently, one finds the corresponding $q(x, t) = Q(k(t, x))$ and $v(t, x) = \frac{q(t, x)}{k(t, x)}$. Hence, the solution $k(t, x)$, $q(t, x)$, and $v(t, x)$ of any time-space point (t, x) can be determined.

Note that the conservation law (and consequently the LWR model) involves three dependent variables: flow (flux) q , density (concentration) k , and speed v . One might be curious about why density k is always chosen as the target variable to work on. This is because density k is unique in that,

by knowing k , one is able to unambiguously determine flow q and speed v on the basis of equilibrium traffic flow models, while flow q and speed u do not have such a property. Readers should be cautioned again that equilibrium traffic flow models are only of statistical significance, and their use as a lookup table is the last resort when no better choice is available.

8.2 EXAMPLE: LWR WITH GREENSHIELDS MODEL

The Greenshields model [9] assumes the following linear v - k relationship:

$$v = v_f \left(1 - \frac{k}{k_j}\right),$$

where v_f is free-flow speed and k_j is jam density. This model implies the following quadratic q - k relationship:

$$q = Q(k) = v_f \left(k - \frac{k^2}{k_j}\right).$$

Hence

$$c(k) = Q'(k) = v_f - 2 \frac{v_f}{k_j} k.$$

If the parameters are traffic speed $v_f = 60$ miles per hour and density $k_j = 240$ vehicles per mile, the explicit form of the LWR model becomes

$$k_t + \left(60 - \frac{k}{2}\right)k_x = 0.$$

Find solutions at points ($t = \frac{1}{2}$ h, $x = 25$ miles) and ($t = 1$ h, $x = 65$ miles) with use of the following initial condition:

$$k(0, x) = k_0(x) \begin{cases} 40 \text{ vehicles per mile} & \text{if } 0 < x \leq 10 \text{ miles,} \\ 20 \text{ vehicles per mile} & \text{if } x > 10 \text{ miles.} \end{cases}$$

Following the above solution procedure, one constructs a time-space diagram, shows the initial condition at the side of the diagram, and identifies the two points in question (see Figure 8.1). Next, one constructs characteristics. All characteristics drawn between $0 < x \leq 10$ miles will bear a k value of 40 vehicles per mile, which can be read from the initial condition, so the slope of these characteristics is $c = 60 - \frac{k}{2} = 40$ miles

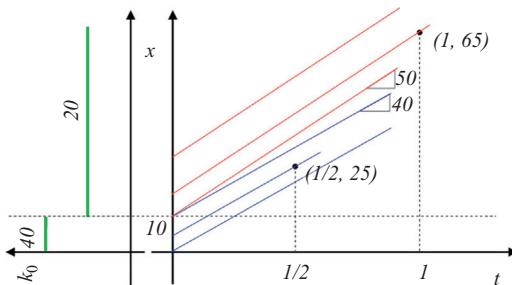


Figure 8.1 Example: LWR with Greenshields model.

per hour. Point $(t = \frac{1}{2}, x = 25)$ is within this area, and the characteristic passing this point intercepts the x -axis at $(0, 5)$. Hence, $k(\frac{1}{2}, 25) = k(0, 5) = 40$ vehicles per mile. Similarly, All characteristics drawn from $x > 10$ miles have slope $c = 50$ miles per hour, and point $(t = 1, x = 65)$ is within this area. The characteristic passing this point intercepts the x -axis at $(0, 15)$. Hence, $k(1, 65) = k(0, 15) = 20$ vehicles per mile.

8.3 SHOCK WAVE SOLUTION TO THE LWR MODEL

The above example actually involves two platoons: a fast one running in front and a slow one trailing behind. Each platoon corresponds to a family of characteristics called a *kinematic wave*. The characteristics of the fast platoon have a slope of 50 miles per hour, which is the speed of the fast kinematic wave. Similarly, the speed of the slow kinematic wave is 40 miles per hour. Noticeably, there is a wedge between the two families of characteristics starting from $(0, 10)$, meaning there is an increasing “vacuum” (or gap) between the two platoons.

If the two platoons are reversed—that is, the slow platoon leads the fast platoon, sooner or later the fast platoon will catch up with the slow platoon. When this occurs, the first vehicle in the fast platoon will have to adopt the speed of the last vehicle in the slow platoon. Shortly afterward, the second vehicle in the fast platoon will have to slow down, and so will the third vehicle, the fourth vehicle, and so on. The “slowing down” effect will propagate backward along the fast platoon. The propagation of a sudden change of traffic condition (e.g., speed drop in this example) creates a *shock wave* which delineates regions of different traffic conditions (e.g., slow and fast traffic in this example). The trajectory of the shock wave in the $x-t$ plane is called a *shock path*.

As discussed in the method of characteristics, a characteristic carries a constant k value (i.e., density), and the intersection of two characteristics will inevitably have two k values. This means that at this point two traffic conditions coexist, and after the intersection, the two platoons resume their original conditions along their respective characteristics. This situation does not make any physical sense. To develop a solution that is physically meaningful, one has to make the solution piecewise smooth. This requires that a characteristic carries one and only one traffic condition (e.g., a k value). When two characteristics meet, both characteristics terminate, and there is a jump (or shock) at the intersection.

To illustrate the idea, the previous example is revisited with the fast platoon being behind. In the $x-t$ plane in Figure 8.2, two families of characteristics—that is, two kinematic waves—are drawn, but this time those characteristics drawn between $0 < x < 10$ will have a slope of 50, while those drawn from $x > 10$ have a slope of 40. Since the fast kinematic wave is behind, it will catch up with the slow kinematic wave—that is, the two families of characteristics will intersect. Whenever two characteristics intersect, they terminate at their intersection. A curve that connects these intersections gives a shock path, along which two regions are delineated: one region belongs to the slow platoon—that is, all points in this region carry the condition of the slow platoon—and the other region belongs to the fast platoon—that is, all points in this region carry the condition of the fast platoon. When one moves across the shock path, the traffic condition changes suddenly from one condition to another—that is, experiencing a shock, which is how a shock wave gets its name. Therefore, it is convenient to read from Figure 8.2 that $k(\frac{1}{2}, 25) = k(0, 0) = 20$ vehicles per mile and $k(1, 65) = k(0, 25) = 40$ vehicles per mile.

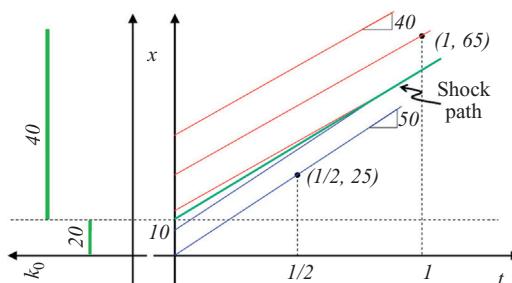


Figure 8.2 Example: LWR with Greenshields model revisited.

8.4 RIEMANN PROBLEM

In the above example, two properties are noticeable:

1. Each of the two kinematic waves consists of a family of straight, parallel characteristics.
2. The shock path is a straight line.

As discussed in Chapter 6, if c is a constant or dependent on k but not explicitly dependent on t or x , the resultant characteristic is a straight line, which is the case in the above example since $c = Q'(k) = v_f - 2\frac{v_c}{k_j}k$.

From Chapter 6, the method of characteristics stipulates that the slope of a characteristic be $\frac{dx}{dt} = c$, which depends on the initial condition. If the initial condition consists of piecewise constant k_0 , each family of characteristics will have the same slope—that is, they are parallel. The above example is such a case.

As discussed in Chapter 7, the slope of a shock path is determined by the Rankine-Hugoniot jump condition. If the initial condition consists of piecewise constant k_0 , then solutions k and q on both sides of the shock path are piecewise constant. Consequently, the Rankine-Hugoniot jump condition will result in a shock path with a constant slope—that is, a straight line—which is also the case in the above example.

Hence, it becomes clear that the solution to an LWR model will always have the above two properties as long as the initial data are given as piecewise constant. In general, a conservation law with piecewise constant initial data is referred to as a *Riemann problem*, named after Bernhard Riemann, who was a German mathematician.

8.5 LWR MODEL WITH A GENERAL Q-K RELATIONSHIP

In the above examples, the underlying q - k relationship is explicitly given—for example, the Greenshields model. Hence, it is convenient to determine the speed of a kinematic wave (i.e., the slope of a family of straight, parallel characteristics) from the initial condition. However, it is recognized that the Greenshields model suffers from inaccuracy, and often the underlying q - k relationship is graphically given by fitting from empirical data. In this case, the solution to the LWR model with a general q - k relationship is typically determined graphically.

Consider the following LWR model with a general q - k relationship:

$$\begin{cases} k_t + q_x = 0, \\ q = Q(k), \\ k(t, 0) = k_0(x) = \begin{cases} A & \text{if } x \leq 0, \\ B & \text{if } x > 0, \end{cases} \end{cases} \quad (8.3)$$

where the underlying q - k relationship is given in Figure 8.3, where A denotes an operating point characterized by flow q_A , density k_A , and speed v_A , and similar notation applies to point B . A time-space diagram is constructed below the q - k relationship with the initial condition at the side. Since this is a Riemann problem, each kinematic wave has a constant slope, and the shock path will be a straight line. From the initial condition, there are two kinematic waves: kinematic wave A emitted from $x \leq 0$, and kinematic wave B emitted from $x > 0$. The speed of kinematic wave A is

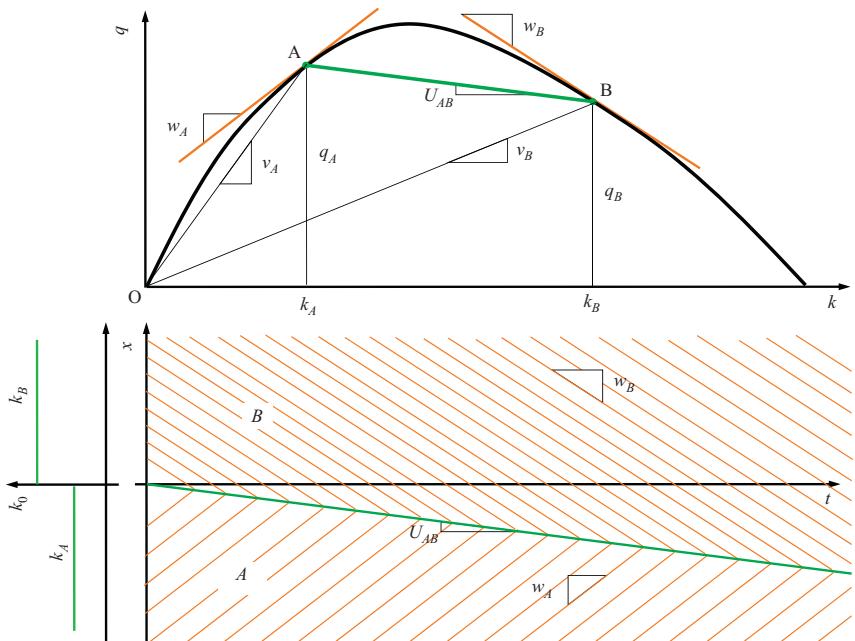


Figure 8.3 Example: LWR model with a general q - k relationship.

$$w_A = \frac{dQ}{dk} = Q'(k)|_{k=k_A};$$

that is, the derivative of the q - k relationship evaluated at operating point A . This is the tangent to the q - k curve at point A . Therefore, one constructs kinematic wave A by drawing a family of straight, parallel lines drawn from $x \leq 0$ with slope w_A . Similarly, the speed of kinematic wave B, w_B , is the tangent to the q - k curve at point B , and the wave can be constructed accordingly. Since kinematic wave B represents a heavy, slow platoon in front and kinematic wave A represents a light, fast platoon behind, kinematic wave A will catch up with kinematic wave B, creating a shock wave. Again, since this is a Riemann problem, the shock path is a straight line. The slope of this line (i.e., the speed of the shock wave) is determined by the Rankine-Hugoniot jump condition:

$$U_{AB} = \frac{q_B - q_A}{k_B - k_A}.$$

This happens to be the slope of the chord connecting points A and B in the q - k curve. In addition, one already knows from the initial condition that the shock path starts at the origin in the time-space diagram. Therefore, one can determine the shock path by drawing a line from the origin with slope U_{AB} . Characteristics in the two kinematic waves will terminate once they meet the shock path. Hence, the shock wave solution is graphically constructed, and consists of two piecewise smooth solutions: the region above the shock path has a uniform traffic condition B (q_B, k_B, u_B) and the region below the shock path has condition A (q_A, k_A, u_A).

8.6 SHOCK PATH AND QUEUE TAIL

In [Figure 8.3](#), the shock path actually represents the time-varying location which separates the fast platoon and the slow platoon—that is, the tail of a moving queue. As the leading vehicle of the fast platoon catches up with the tail of the slow platoon, that vehicle joins the slow platoon and becomes its new tail. Since the slow platoons is still moving, the location of its tail changes dynamically depending on how quick the fast platoon arrives. [Figure 8.4](#) shows a few snapshots to illustrate such a dynamic process.

One may have recognized that although characteristics are used to illustrate how to find the shock path, they are actually unnecessary. With a known point on the shock path and known shock speed, the shock path can be determined directly without characteristics being drawn. In the above

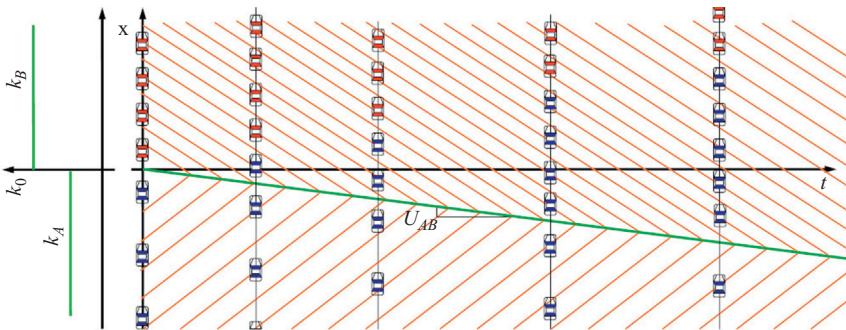


Figure 8.4 Shock path and queue tail.

example, one can construct the solution directly by drawing a line from the origin with slope U_{AB} . This line is the shock path and also the queue tail which separates regions with conditions A and B.

8.7 PROPERTIES OF THE FLOW-DENSITY RELATIONSHIP

It can be seen from the above example that the flow-density ($q-k$) relationship is very illustrative to show various speeds. [Figure 8.5](#) gives the full picture.

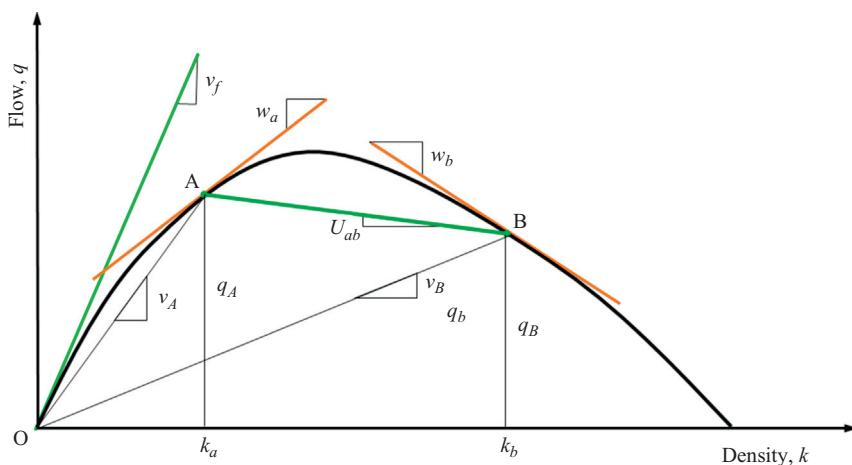


Figure 8.5 Speeds in a flow-density relationship.

8.7.1 Flow-Density Relationship and Speeds

Traffic speed v

If operating point A , which represents a traffic condition with flow q_A and density k_A , is known, the corresponding traffic speed under condition A, by definition, is

$$v_A = \frac{q_A}{k_A}.$$

Graphically, this can be represented as the slope of the line connecting the origin O and operating point A .

Free-flow speed v_f

If k_A decreases, point A will move along the curve toward the origin O . In the limiting case where $k_A \rightarrow 0$, line OA becomes the tangent to the curve at the origin. The slope of this tangent denotes the traffic speed when the density is close to zero. By definition, the slope represents the free-flow speed v_f :

$$v_f = \lim_{A \rightarrow O} v_A = \lim_{k_A \rightarrow 0} \frac{q_A}{k_A}.$$

Kinematic wave speed w

If one draws a line tangent to the curve at point A , as discussed above, the slope of this tangent is the speed of a kinematic wave carrying traffic condition A:

$$w_A = Q'(k)|_{k=k_A}.$$

Shock wave speed U

If A and B represent two different traffic conditions, as discussed above, the slope of chord AB is the speed of the shock wave should traffic with condition A catch up with traffic with condition B.

$$U_{AB} = \frac{q_B - q_A}{k_B - k_A}.$$

8.7.2 Flow-Density Relationship Observed by a Moving Observer

Note that the above discussion is based on the observation from the perspective of a stationary observer—that is, everything is relative to an

observer standing stationary at the roadside. Now what happens if the observer is moving? For example, what if the observer is riding on the kinematic wave carrying traffic condition A? The moving observer will now observe less flow than he or she would have observed if he or she were stationary. The following equation quantifies the relative flow that the moving observer sees:

$$\tilde{q}_A = q_A - w_A k_A.$$

This is equivalent to drawing a line from the origin with w_A as its slope. Then run a vertical line through point A intersecting the drawn line at A'' and the horizontal axis at A' . The length of AA' is q_A , the segment of $A'A''$ is $w_A k_A$, and the segment of AA'' is the relative flow, \tilde{q}_A , observed by the moving observer. As another example, suppose traffic is operating at condition B which is on the congested side of the q - k curve. The kinematic wave speed is now w_B , which is negative. What happens if an observer is moving along with wave w_B ? With the same treatment, one obtains

$$\tilde{q}_B = q_B - w_B k_B.$$

This is equivalent to drawing a line from the origin O with slope w_B which slants downward. Run a vertical line through point B intersecting the drawn line at B'' and the horizontal axis at B' . The absolute value of relative flow (i.e., the length of BB'') in this case is the sum of BB' and $B'B''$ because w_B takes a negative value.

8.8 EXAMPLE LWR MODEL PROBLEMS

The above discussion focused on the LWR model and its solutions using the method of characteristics and shock waves. It is time to apply this method to solve some concrete traffic flow problems.

8.8.1 A Bottleneck with Varying Traffic Demand

Traffic arriving at the upstream point of a highway was initially under condition A (see [Table 8.1](#) and [Figure 8.7](#)). At 9:00 a.m., the arriving traffic switches to condition B. After 1 h, the arriving traffic switches back to condition A. The capacity at the bottleneck is 1400 vehicles per hour. Find how far the queue extends back and how long the queue persists.

Table 8.1 Traffic data: a bottleneck with varying traffic demand

Condition	q (vehicles/h)	k (vehicles/km)	v (km/h)
A	600	8.57	70
B	2000	40	50
D	1400	21.5	65
D'	1400	130	10.8

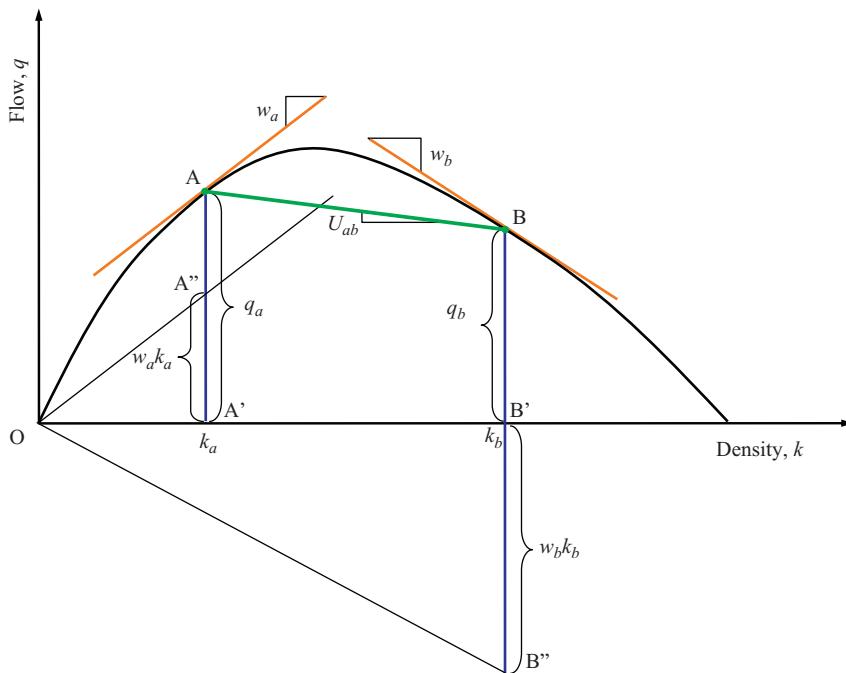


Figure 8.6 Traffic flow observed by a moving observer.

Solution. With the aid of the graphical construction in Figure 8.7, the rate at which the queue grows is:

$$U_{BD'} = \frac{q_{D'} - q_B}{k_{D'} - k_B} = \frac{1400 - 2000}{130 - 40} = -\frac{600}{90} = -6.67 \text{ km/h.}$$

The queue tail extends back at this rate for 1 h, so the farthest point it reaches is 6.67 km upstream of the bottleneck. The rate at which the queue dissipates is

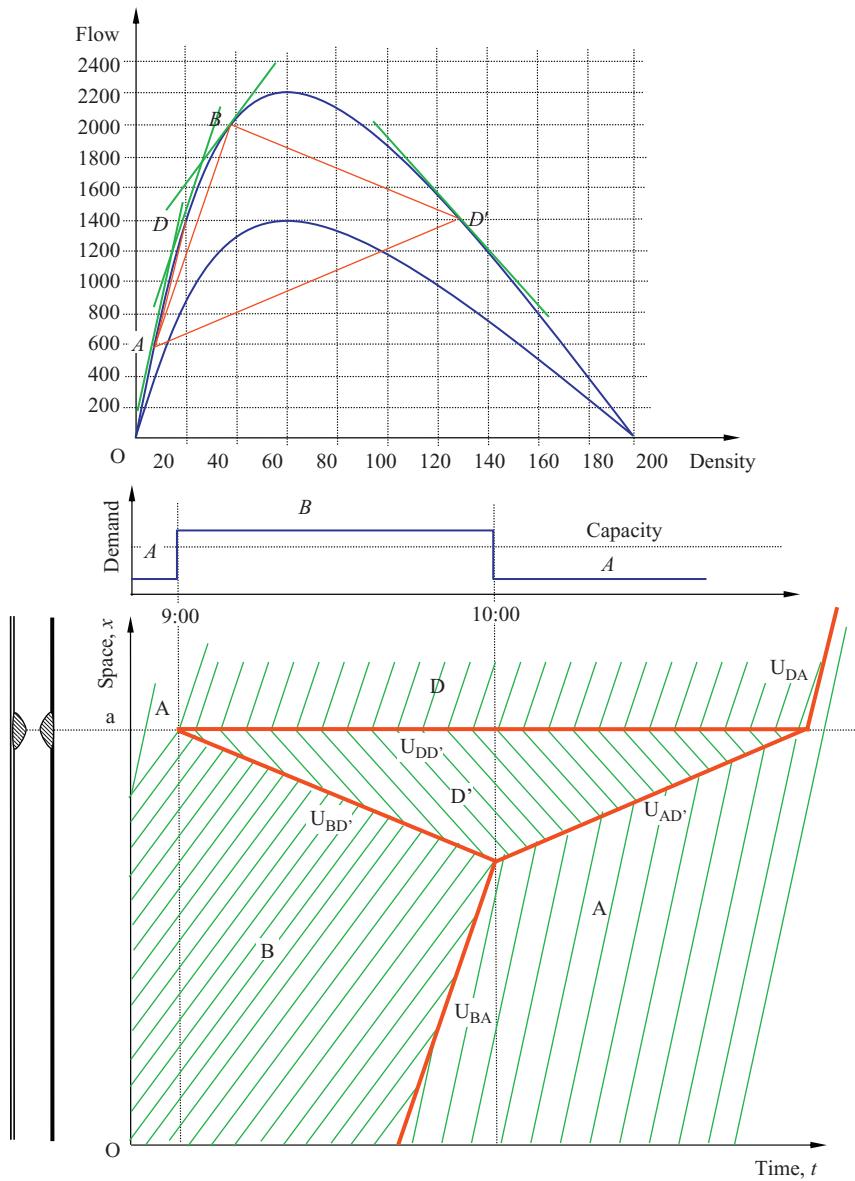


Figure 8.7 A highway bottleneck with varying traffic demand.

Table 8.2 Traffic data: a moving bottleneck

Condition	q (vehicles/h)	k (vehicles/km)	v (km/h)
A	700	10	70
B	1600	120	13.3
C	2200	60	36.7
O	0	0	75

$$U_{AD'} = \frac{q_{D'} - q_A}{k_{D'} - k_A} = \frac{1400 - 600}{130 - 8.57} = 6.60 \text{ km/h.}$$

So the time needed to dissipate the queue is $\frac{6.67}{6.60} = 1.01$ h, and the total time for which the queue persists is 2.01 h.

8.8.2 A Moving Bottleneck

A freeway was initially operating under condition A (see [Table 8.2](#)). At 2:30 p.m., a sluggish truck entered the freeway traveling at a speed of 13.3 km/h. The truck turned off the freeway at the next exit 6.67 km away. Find when the impact of the truck will disappear.

Solution. With the aid of the graphical construction in [Figure 8.8](#), the following can be calculated:

$$U_{OB} = \frac{q_B - q_O}{k_B - k_O} = \frac{1600 - 0}{120 - 0} = 13.30 \text{ km/h,}$$

$$U_{AB} = \frac{q_B - q_A}{k_B - k_A} = \frac{1600 - 700}{120 - 10} = 8.18 \text{ km/h,}$$

$$U_{CB} = \frac{q_B - q_C}{k_B - k_C} = \frac{1600 - 2200}{120 - 60} = -10.00 \text{ km/h,}$$

$$\frac{be}{ae} = U_{OB} \rightarrow ae = \frac{be}{U_{OB}} = \frac{6.67}{13.3} = 0.5 \text{ h,}$$

$$\frac{cd}{bc} = U_{CB} \rightarrow cd = U_{CB} \times bc = 10bc,$$

$$\frac{df}{af} = U_{AB} \rightarrow df = U_{AB} \times af = 8.18af,$$

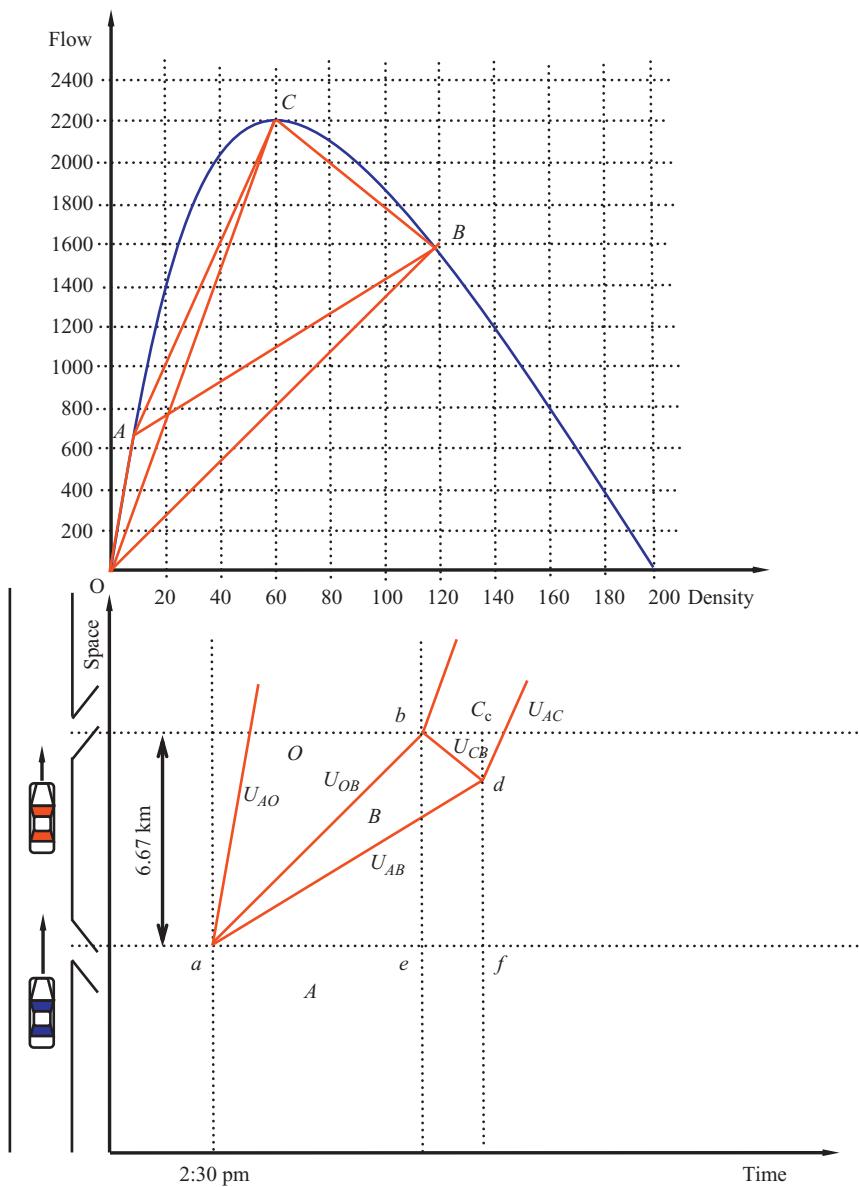
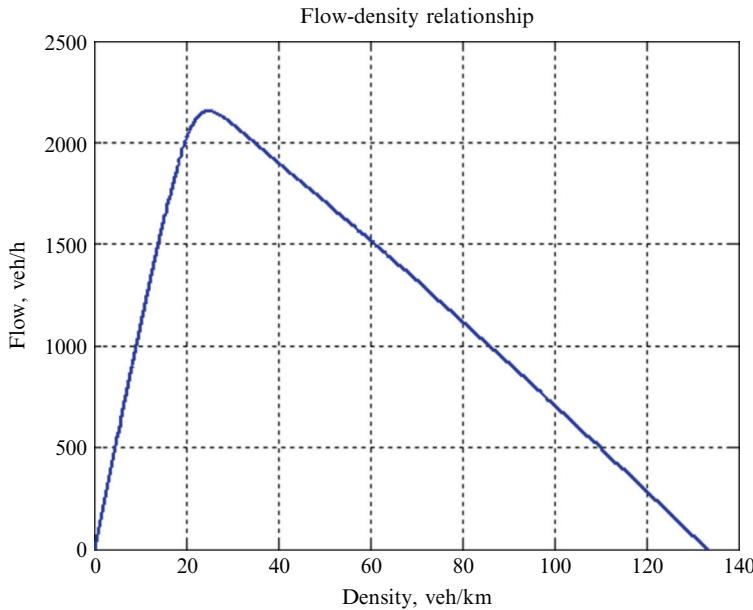


Figure 8.8 A moving bottleneck with constant demand.



$$\begin{cases} 10bc + 8.18af = 6.67, \\ af - bc = 0.5, \end{cases}$$

$$af = 0.64 \text{ h.}$$

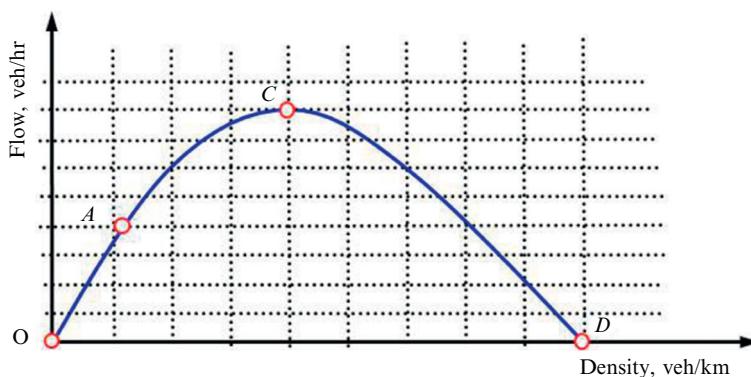
So the impact of the truck lasts for 0.64 h.

PROBLEMS

1. The figure below illustrates a hypothetical flow-density relationship. Identify the following from the figure (mark them on the figure if necessary to show your answers):
 - a. Free-flow speed v_f
 - b. Jam density k_j
 - c. Capacity condition (q_m, k_m, v_m)
 - d. Traffic condition at point A (q_A, k_A, v_A)
 - e. Traffic condition at point B (q_B, k_B, v_B)
 - f. Kinematic wave speeds (i) when the density is zero, (ii) at jam density, (iii) at capacity, (iv) at condition A, and (v) at condition B
 - g. Shock wave speed when a platoon of vehicles at condition A catches up with a platoon of vehicles at condition B

2. Assume that traffic on a uniform freeway section follows the above flow-density relationship. At one time, traffic is operating at condition A.
 - a. Find the relative flow observed by a moving observer who is traveling with the traffic at 25 km/h (assume no interaction between the observer and the traffic, e.g., the observer flies over the traffic).
 - b. Similarly, find the relative flow when the traffic condition is B and the observer is riding on the kinematic wave carrying condition B.
3. Traffic on a 16 km uniform segment of Interstate 90 was initially operating at condition B, as illustrated in the figure for problem 1. Starting at 7:00 p.m. and upstream of the midpoint of the uniform section, demand drops and traffic begins to operate at condition A. Assume that the flow-density relationship in the figure applies.
 - a. Determine the traffic condition at a location 2 km downstream of the midpoint at 7:30 p.m.
 - b. Determine the traffic condition at a location 2 km upstream of the midpoint at 7:30 p.m.
 - c. When will the end of the queue reach the upstream end of the uniform section?
4. Traffic on a 16 km uniform segment of Interstate 90 was initially operating at condition A, as illustrated in the figure for problem 1. Starting at 7:00 a.m. and upstream of the midpoint of the uniform section, demand increases and traffic begins to operate at condition B. Assume that the flow-density relationship in the figure applies.
 - a. Determine the traffic condition at a location 2 km downstream of the midpoint at 7:30 a.m.
 - b. Determine the traffic condition at a location 2 km upstream of the midpoint at 7:30 a.m.
5. An intersection with constant demand. Traffic arrives at an approach of a signalized intersection at a constant rate of 800 vehicles per hour. All conditions are given in the table and the flow-density relationship below. The intersection is under pretimed signal control with a cycle length of 90 s and a split of effective green/red of 0.5/0.5. Determine the farthest point of the queue.
6. An intelligent transportation system problem. On Wednesday at 9:00 a.m., there is an accident on northbound Interstate 91. The traffic operation center (TOC) has to decide how to clean up the accident. After collecting information and communicating with highway patrol and emergency operator, the TOC determines that there are two alternatives:

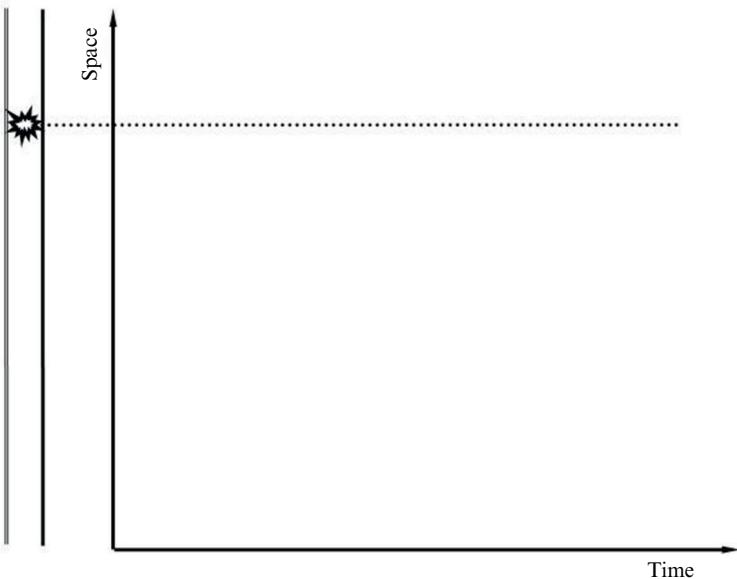
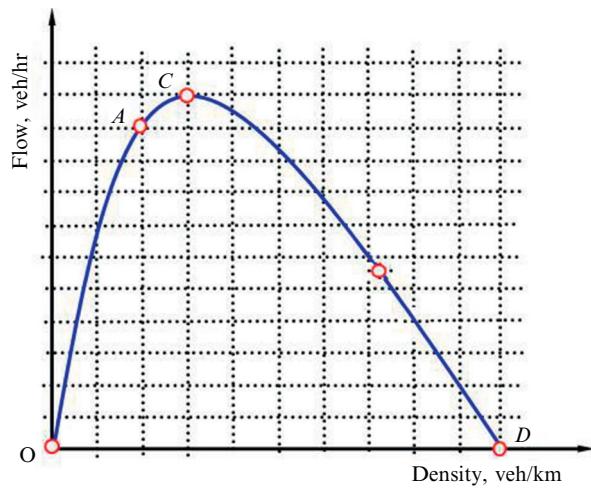
Condition	q (vehicles/h)	k (vehicles/km)	v (km/h)
A	800	25	32
C	1600	80	20
D	0	180	0
O	0	0	40



- a. Completely shut Interstate 91 for 10 min, clean it up, and then reopen Interstate 91 for normal operation.
- b. Partially open Interstate 91 at reduced capacity, but the cleanup requires longer—about 30 min—before normal operation can resume.

One of the concerns at the TOC is how far the queue will spill back because the queue on Interstate 91 will overflow via ramps and further block upstream surface streets. As a transportation engineering student, you are asked to offer your knowledge to help the TOC make a decision. More details are given in the table and the flow-density relationship below. Find which alternative creates the longer queue.

Condition	Description	q (vehicles/h)	k (vehicles/km)	v (km/h)
A	Arrival flow	2000	40	50
D	Queued flow	0	200	0
C	Capacity flow	2200	60	36.7
E	Reduced capacity flow	1100	50	22



CHAPTER 9

Numerical Solutions

Chapter 8 presented the LWR model and the procedure for its solution. In addition, a few concrete examples were provided to show how to apply the procedure. These problems were solved graphically by manually working on a time-space diagram using the method of characteristics. Though illustrative, the graphical approach has limitations since it is capable of dealing only with simple problems which involve only one homogeneous highway section and simple initial conditions. In the real world, a traffic system may consist of a network where multiple segments (links) or highways are considered with traffic flowing in and out via ramps. In addition, the initial and boundary conditions may be more complicated and time-varying. In these cases, the graphical approach is insufficient and sometimes infeasible. Moreover, the purpose of solving LWR problems is to predict traffic dynamics so that traffic engineers are able to anticipate congestion and to develop strategies to alleviate congestion. In such applications, timing is a critical issue, and solving these problems in real time is desirable. Moreover, the wide deployment of intelligent transportation systems makes it possible to provide real-time traffic conditions and allow online prediction. Therefore, a computerized solution to the LWR model is essential to cope with more complicated real-world problems, to enable real-time prediction, and to automate such predictions by the development of online applications.

9.1 DISCRETIZATION SCHEME

Computers are digital machines which can work only in a discrete fashion, so computerized solutions to the LWR model have to be numerical and discrete. The first step to develop a computerized solution is to discretize time and space. Figure 9.1 illustrates a time-space diagram where time t is the horizontal axis and space x is the vertical axis with a roadway drawn at the side. The roadway is partitioned into a series of segments labeled as $j \in (0, 1, \dots, J)$. If x_0 is chosen as the reference point and segment length Δx is uniform, the location of the end of segment j is

$$x_j = x_0 + j\Delta x.$$

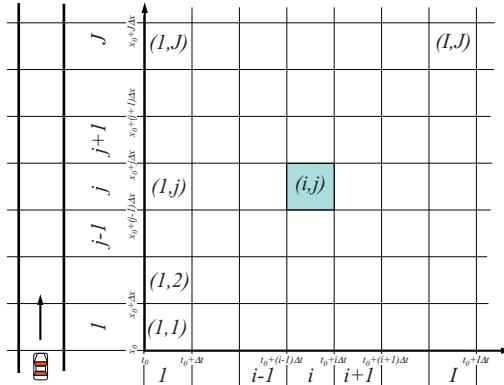


Figure 9.1 Discretization scheme.

Similarly, the time is divided into a series of durations $i \in (0, 1, \dots, I)$ with step size Δt . If the reference point of time is t_0 , the end of duration i is at time

$$t_i = t_0 + i\Delta t.$$

In general, the following relationship is required in a discretization scheme:

$$\frac{\Delta x}{\Delta t} > v_f,$$

where v_f is the free-flow speed. This requirement basically says that a vehicle should not traverse more than one segment Δx within a time step Δt .

A typical numerical solution to the LWR problem starts with initial conditions by determining the number of vehicles contained in each roadway segment one by one from the upstream end to the downstream end:

```

when i = 1
determine storage in j = 1
determine storage in j = 2
...
determine storage in j = J
end

```

For easy reference, the time-space region bounded within duration i and segment j is referred to as a *cell* and is denoted as (i, j) and the number of vehicles contained in segment j at the end of duration i is denoted as $n(t_i, x_j)$. The above listing can be rewritten as

```

when i = 1
    determine n(t_1,x_1)
    determine n(t_1,x_2)
    ...
    determine n(t_1,x_J)
end

```

After this, time advances one step, and the above process starts over again.

```

when i = 2
    determine n(t_2,x_1)
    determine n(t_2,x_2)
    ...
    determine n(t_2,x_J)
end

```

Hence, the numerical solution consists of two loops: time t_i as the outer loop and space x_j as the inner loop:

Numerical solution procedure:

```

for i = 1 to I
    for j = 1 to J
        determine n(t_i,x_j)
    end
end

```

The process finishes when all cells have been traversed, and the solution is given as cell storage $[n(t_i, x_j) | i \in (1, 2, \dots, I), j \in (1, 2, \dots, J)]$ or, alternatively, traffic condition $k(t_i, x_j)$, $q(t_i, x_j)$, and $v(t_i, x_j)$.

Building on the above procedure, we discuss a few numerical solutions to traffic dynamic problems in the following subsections.

9.2 FREFLO

FREFLO is an early (if not the earliest) computerized macroscopic traffic simulation model, developed by Payne [21] in the late 1970s. Like the LWR model, FREFLO consists of three equations with a discretization scheme, shown in Figure 9.2. The first equation is the conservation law:

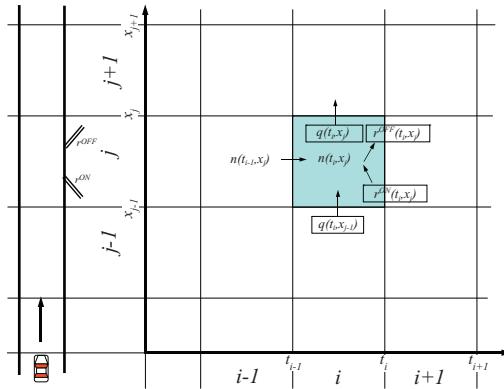


Figure 9.2 Discretization in FREFLO.

storage in the current cell = storage at previous step +
vehicles arrived from upstream - vehicles departed to downstream +
vehicles entered via on-ramp - vehicles exited via off-ramp

Mathematically, this can be expressed as

$$n(t_i, x_j) = n(t_{i-1}, x_j) + \Delta t q(t_i, x_{j-1}) - \Delta t q(t_i, x_j) + \Delta t g(t_i, x_j),$$

where $g(t_i, x_j)$ is the net inflow via ramps—that is, $g(t_i, x_j) = r^{on}(t_i, x_j) - r^{off}(t_i, x_j)$. Note that $n = k\Delta x$, and the above equation becomes

$$k(t_i, x_j)\Delta x = k(t_{i-1}, x_j)\Delta x + \Delta t q(t_i, x_{j-1}) - \Delta t q(t_i, x_j) + \Delta t g(t_i, x_j),$$

$$k(t_i, x_j) = k(t_{i-1}, x_j) + \frac{\Delta t}{\Delta x} [q(t_i, x_{j-1}) - q(t_i, x_j) + g(t_i, x_j)].$$

The second equation of FREFLO is the identity in discrete form:

$$q(t_i, x_j) = k(t_i, x_j)v(t_i, x_j).$$

Different from most first-order models, which adopt an equilibrium speed-density relationship, FREFLO uses a dynamic speed-density relationship as the third equation:

speed in current cell = speed in previous step - convection +
relaxation + anticipation

where:

convection - vehicles tend to continue their speeds when they travel in the upstream section,

relaxation - vehicles tend to adopt the equilibrium velocity-density relationship,

anticipation - vehicles tend to adjust to downstream condition, i.e. slow down if congested.

Mathematically, this can be expressed as

$$\begin{aligned} v(t_i, x_j) = & v(t_{i-1}, x_j) - \Delta t \{ v(t_{i-1}, x_j) \frac{v(t_{i-1}, x_j) - v(t_{i-1}, x_{j-1})}{\Delta x_i} \\ & + \frac{1}{T_j} [v(t_{i-1}, x_j) - V(k(t_{i-1}, x_j))] \\ & + \frac{b_j}{k(t_{i-1}, x_j)} \frac{k(t_{i-1}, x_{j+1}) - k(t_{i-1}, x_j)}{\Delta x_j} \}], \end{aligned}$$

where $T_j = c_T \Delta x_j$ and $b_j = q_b \Delta x_j$. c_T and q_b are relaxation time and anticipation coefficients, respectively. The equilibrium speed-density relationship $V(k)$ takes the following form:

$$v = V(k) = \min\{88.5, (172 - 3.72k + 0.0346k^2 - 0.00119k^3)\},$$

which was an empirical speed-density relationship obtained by least-squares fitting of observed data.

With the above equations, one is able to determine the state (q, k, v) of each cell by starting from initial conditions and following the numerical solution procedure.

9.3 FREQ

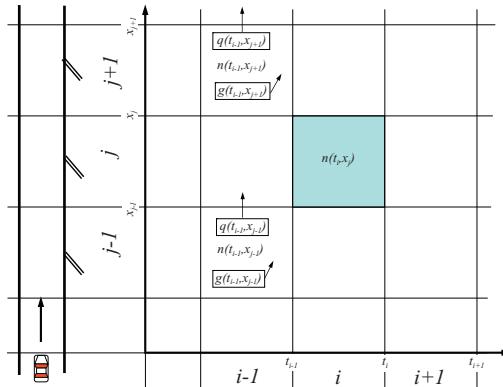
FREQ is a computerized macroscopic traffic simulation model developed by May [26] in the early 1980s. Its underlying algorithm is not publicly available.

9.4 KRONOS

KRONOS is another computerized macroscopic traffic simulation model, developed by Michalopoulos [27] in the mid-1980s. In addition to proposing a numerical solution to the LWR model, Michalopoulos enriched the solution by incorporating ramp flows and lane changes. If net ramp flow $g(t, x)$ is considered, the continuity equation becomes:

$$k_t + q_x = g(t, x).$$

The discrete form of the equation can be stated as

**Figure 9.3** Discretization in KRONOS.

Storage in the current cell =

Average of storages in upstream and downstream segments at previous step -

Average of mainline net outflows in upstream and downstream segments at previous step +

Average of ramp net inflows in upstream and downstream segments at previous step

See the illustration in [Figure 9.3](#). Mathematically, this is equivalent to

$$n(t_i, x_j) = \frac{n(t_{i-1}, x_{j+1}) + n(t_{i-1}, x_{j-1})}{2} - \frac{\Delta tq(t_{i-1}, x_{j+1}) - \Delta tq(t_{i-1}, x_{j-1})}{2} + \frac{\Delta tg(t_{i-1}, x_{j+1}) + \Delta tg(t_{i-1}, x_{j-1})}{2}.$$

Note that $n = \Delta x k$, and the above equation becomes

$$\Delta x k(t_i, x_j) = \frac{\Delta x k(t_{i-1}, x_{j+1}) + \Delta x k(t_{i-1}, x_{j-1})}{2} - \frac{\Delta tq(t_{i-1}, x_{j+1}) - \Delta tq(t_{i-1}, x_{j-1})}{2} + \frac{\Delta tg(t_{i-1}, x_{j+1}) + \Delta tg(t_{i-1}, x_{j-1})}{2}.$$

Hence,

$$\begin{aligned} k(t_i, x_j) = & \frac{k(t_{i-1}, x_{j+1}) + k(t_{i-1}, x_{j-1})}{2} \\ & - \frac{\Delta t}{\Delta x} \frac{q(t_{i-1}, x_{j+1}) - q(t_{i-1}, x_{j-1})}{2} \\ & + \frac{\Delta t}{\Delta x} \frac{g(t_{i-1}, x_{j+1}) + g(t_{i-1}, x_{j-1})}{2}. \end{aligned}$$

This equation is supplemented by the identity

$$q(t_i, x_j) = k(t_i, x_j)v(t_i, x_j)$$

and an equilibrium relationship

$$v(t_i, x_j) = V(k(t_i, x_j)),$$

the simplest form of which is the Greenshields model [9]:

$$v(t_i, x_j) = v_f(1 - k(t_i, x_j)/k_j)$$

The initial condition is given as at t_0 : k , q , u is known at all locations x_j , $j = 0, 1, 2, \dots, J$.

The boundary condition is given as $q(t_i, x_0)$, $i = 0, 1, 2, \dots, I$, and $g(t_i, x_J)$, $i = 0, 1, 2, \dots, I$ and $j = 0, 1, 2, \dots, J$

Starting from the initial condition and applying the boundary condition, one can work out the numerical solution by following the numerical solution procedure.

9.5 CELL TRANSMISSION MODEL

The cell transmission model (CTM) was proposed by Daganzo [28, 29] in the mid-1990s. The model was presented in two papers, with the first addressing mainline traffic and the second addressing network traffic.

9.5.1 Minimum Principle

Figure 9.4 shows a triangular flow-density relationship. The relationship consists of three sections: uncongested (left), with free-flow speed v_f equal to forward wave kinematic speed w_f , capacity (middle) q_m , and congested (right), with backward wave speed w_b and jam density K .

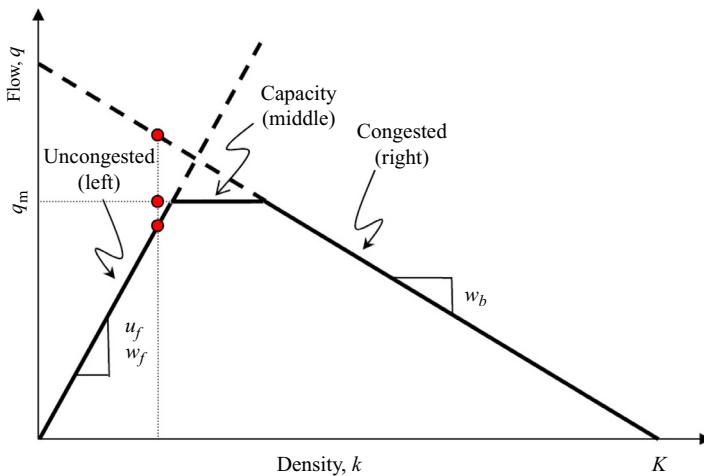


Figure 9.4 Triangular flow-density relationship.

A vertical line at any density k will intersect the three sections at height kw_f , q_m , and $(K - k)w_b$. Hence, flow corresponding to this density is found as the minimum of the three intersections:

$$q = \min\{kw_f, q_m, (K - k)w_b\}$$

Physically, if one considers the left section as conditions dictated by arrival traffic, the middle section as local capacity, and the right section as conditions dictated by downstream traffic, the above equation basically says that traffic flowing through a point of highway should not exceed the upstream arrival rate, local capacity, and the rate allowed by downstream conditions.

9.5.2 Mainline Scenario

The CTM uses the same discretization scheme presented in [Figure 9.1](#). Everything else remains the same except for one thing: the cell now has a uniform length as the distance traveled by a vehicle at free-flow speed during one time step:

$$\Delta x = v_f \Delta t$$

According to the minimum principle, traffic that can flow into segment j , $q_j(t_i)$, is constrained by the following:

$$q_j(t_i) = \min\{k_{j-1}(t_{i-1})w_f, q_m, (K - k_j(t_{i-1}))w_b\}.$$

Hence, the number of vehicles that can move into segment j , $\gamma_j(t_i)$, is found by multiplying both sides by Δt :

$$\gamma_j(t_i) = q_j(t_i)\Delta t = \min\{k_{j-1}(t_{i-1})w_f\Delta t, q_m\Delta t, (K - k_j(t_{i-1}))w_b\Delta t\}.$$

Note that $n = k\Delta x$, $\Delta x = v_f\Delta t$, and $v_f = w_f$ owing to the triangular flow-density relationship. The above equation can be transformed to the following form:

$$\gamma_j(t_i) = \min\{k_{j-1}(t_{i-1})\Delta x, q_m\Delta t, \frac{w_b}{w_f}(K - k_j(t_{i-1}))\Delta x\};$$

that is,

$$\gamma_j(t_i) = \min\{n_{j-1}(t_{i-1}), q_m\Delta t, \frac{w_b}{w_f}(K\Delta x - n_j(t_{i-1}))\}.$$

The above equation stipulates that the number of vehicles that can move into segment j , $\gamma_j(t_i)$, is constrained by

- the number of vehicles in $j - 1$ previously: $n_{j-1}(t_{i-1})$,
- the capacity of segment j , $q_m\Delta t$, and
- the empty space in j : $\frac{w_b}{w_f}(K\Delta x - n_j(t_{i-1}))$.

The equation can be further reduced to

$$\gamma_j(t_i) = \min\{S_{j-1}, R_j\},$$

where $S_{j-1} = \min\{n_{j-1}(t_{i-1}), q_m\Delta t\}$ represents flow being sent from an upstream position and $R_j = \min\{q_m\Delta t, \frac{w_b}{w_f}(K\Delta x - n_j(t_{i-1}))\}$ is flow ready to be received downstream.

Therefore, the evolution of traffic on a freeway mainline can be stated as

```

Storage in current cell =
Storage in the cell previously +
Vehicles flowed in -
Vehicles flowed out

```

Mathematically, this can be expressed as

$$n_j(t_i) = n_j(t_{i-1}) + \gamma_j(t_i) - \gamma_{j+1}(t_i).$$

9.5.3 Merger Scenario

To be able to address network traffic, a queuing model is needed for a merger where two streams of traffic flow into one. The merger consists of two upstream links (e.g., a mainline link $(j - 1)$ and an on-ramp $(j - 1)'$) and one downstream link j (see Figure 9.5). Assume that during interval (t_{i-1}, t_i) links $(j - 1)$ and $(j - 1)'$ have S_{j-1} and S'_{j-1} vehicles to send, respectively, and link j can receive R_j vehicles. Considering that demand (i.e., $S_{j-1} + S'_{j-1}$) and supply (i.e., R_j) may not match in this case, link $(j - 1)$ actually sends y_{j-1} vehicles into link j and link $(j - 1)'$ actually sends y'_{j-1} vehicles, where $y_{j-1} \leq S_{j-1}$, $y'_{j-1} \leq S'_{j-1}$, and $y_{j-1} + y'_{j-1} \leq R_j$. In addition, mainline and on-ramp traffic have their relative priorities p_{j-1} and p'_{j-1} , respectively, where $p_{j-1} \geq 0$, $p'_{j-1} \geq 0$, and $p_{j-1} + p'_{j-1} = 1$. The merger queuing model is essentially solving for y_{j-1} and y'_{j-1} given S_{j-1} , S'_{j-1} , R_j , p_{j-1} , and p'_{j-1} .

Figure 9.6 illustrates how to find the solution. The horizontal and vertical axes are y_{j-1} (mainline outflow) and y'_{j-1} (on-ramp outflow), respectively. A rectangle is constructed as being bounded by the two axes, a horizontal line at $y'_{j-1} = S'_{j-1}$ and a vertical line at $y_{j-1} = S_{j-1}$. The latter two intersect at point A . Draw a line from the origin O with slope $\frac{p'_{j-1}}{p_{j-1}}$ and the line intersects the rectangle at point C . Curve ACO denotes the collection of solutions and reason is as follows.

Given the sending flows S_{j-1} and S'_{j-1} and receiving flow R_j , there are three possibilities:

1. *Supply exceeds demand:* This is to say that $S_{j-1} + S'_{j-1} \leq R_j$. Physically, this means that link j is able to receive more vehicles than the total to be sent from both upstream links. For example, $S_{j-1} = 100$, $S'_{j-1} = 80$, and $R_j = 200$. In this case, vehicles from both upstream links can flow into the downstream link without any problem. Graphically, this situation corresponds to a line (e.g., line 1) which represents the collection of

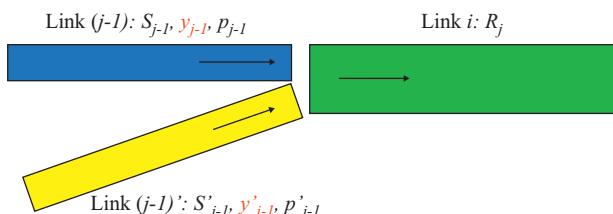


Figure 9.5 A freeway merge.

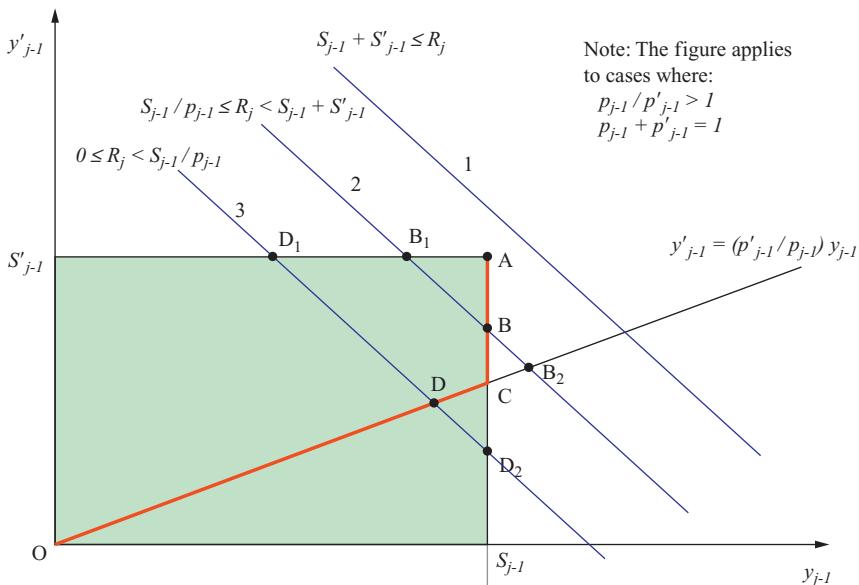


Figure 9.6 Queuing at a freeway merger.

points whose coordinates sum up to R_j . Such a line is always to the right of vertex A without intersecting the rectangle which represents the collection of all feasible solutions. Therefore, the solution is

$$\begin{cases} y_{j-1} = S_{j-1} & \text{if } R_j \geq S_{j-1} + S'_{j-1} \\ y'_{j-1} = S'_{j-1} & \end{cases} \quad (9.1)$$

This solution corresponds to vertex A in Figure 9.6.

2. *Demand exceeds supply and one upstream link is congested:* This is to say that $S_{j-1} + S'_{j-1} > R_j$. In addition, one upstream link fails to send all vehicles that it has. For example, $S_{j-1} = 100$, $S'_{j-1} = 80$, and $R_j = 160$. The priority rules stipulate a split of $\frac{p_{j-1}}{p'_{j-1}} = \frac{3}{1}$, meaning that, among the 160 spaces downstream, the mainline can send $160 \times \frac{3}{4} = 120$ vehicles and the on-ramp can send $160 \times \frac{1}{4} = 40$ vehicles. Since the mainline has only 100 vehicles to send, these vehicles are able to enter link j without delay, leaving $160 - 100 = 60$ spaces in link j for traffic from the on-ramp. The on-ramp has 80 vehicles to send, 60 of which are admitted

by link j and the remaining 20 are delayed. Summing up, the solution to this example is

$$y_{j-1} = S_{j-1} = 100, y'_{j-1} = R_j - S_{j-1} = 160 - 100 = 60.$$

Graphically, this situation corresponds to a line (e.g., line 2) which is parallel to line 1 and intersects the rectangle between points A and C . Line 2 consists of all points whose coordinates sum up to R_j . This line intersects line $y'_{j-1} = S'_{j-1}$ at point B_1 , line $y_{j-1} = S_{j-1}$ at point B , and the priority line $y'_{j-1} = \frac{p'_{j-1}}{p_{j-1}} y_{j-1}$ at point B_2 . The three intersections are three feasible solutions. With sue of the above example as an illustration, these solutions can be interpreted as follows:

Point B_1

Suggests that $S'_{j-1} = 80$ vehicles from the on-ramp can depart without delay and the remaining $160 - 80 = 80$ spaces in link j can be used to admit 80 of the 100 vehicles from link $j - 1$. This violates the priority rule.

Point B

Suggests that $S_{j-1} = 100$ vehicles from link $j - 1$ can depart without delay and the remaining $160 - 100 = 60$ spaces in link j can be used to admit 60 of the 80 vehicles from the on-ramp. This is the correct solution.

Point B_2

Suggests that link j will admit $160 \times \frac{3}{4} = 120$ vehicles from link $j - 1$ and the remaining $160 - 120 = 40$ remaining spaces in link j can be used to admit 40 of the 80 vehicles from the on-ramp. Since more vehicles cannot depart from link $j - 1$ than link $j - 1$ has, this solution is incorrect.

From the outcome of the example, it is clear that the true solution is point B , which is the middle of the three points. Mathematically, this can be expressed as follows:

$$\begin{cases} y_{j-1} = \text{mid}\{S_{j-1}, R_j - S'_{j-1}, p_{j-1}R_j\} & \text{if } R_j < S_{j-1} + S'_{j-1}, \\ y'_{j-1} = \text{mid}\{S'_{j-1}, R_j - S_{j-1}, p'_{j-1}R_j\} & \end{cases} \quad (9.2)$$

where the *mid* operator takes the middle value of all the members. Line segment AC contains all solutions of this nature.

- 3. Demand exceeds supply and both upstream links are congested:** This is to say that $S_{j-1} + S'_{j-1} > R_j$. In addition, both upstream links fail to send all vehicles that they have. We use the above example except that $R_j = 120$. The priority rules stipulate that link $j - 1$ can send as a maximum $120 \times \frac{3}{4} = 90$ vehicles to link j and the on-ramp can send $120 \times \frac{1}{4} = 30$ vehicles. Since both upstream links have more vehicles than they are able to send, the priority rule takes control—that is, link $j - 1$ will actually send 90 vehicles, with the remaining $100 - 90 = 10$ vehicles delayed, and the on-ramp will send 30 vehicles, leaving 50 vehicles delayed.

Graphically, this situation corresponds to a line (e.g., line 3) which is parallel to line 1 and intersects the priority line between points C and O . Again, line 3 consists of all points whose coordinates sum up to R_j . From the above example, line 3 intersects line $y'_{j-1} = S'_{j-1}$ at point D_1 , the priority line $y'_{j-1} = \frac{p'_{j-1}}{p_{j-1}} y_{j-1}$ at point D , and line $y_{j-1} = S_{j-1}$ at point D_2 . The three intersections are three feasible solutions, and their physical meaning is as follows:

Point D_1

Suggests that $S'_{j-1} = 80$ vehicles from the on-ramp can depart without delay and the remaining $120 - 80 = 40$ spaces in link j can be used to admit 40 of 100 vehicles from link $j - 1$. This violates the priority rule.

Point D

Follows the priority rule by allowing 90 of the 100 vehicles from link $j - 1$ to enter link j and using the remaining $120 - 90 = 30$ spaces to admit 30 of the 80 vehicles from the on-ramp. This is the correct solution.

Point D_2

Suggests that link $j - 1$ can actually send $S_{j-1} = 100$ vehicles to link j and the remaining $120 - 100 = 20$ spaces are used to admit 20 of the 80 vehicles from the on-ramp. This, again, violates the priority rule.

Therefore, the true solution is, again, the middle of the three points, and the mathematical notation is the same as above. In addition, line segment CO contains all solutions of this nature.

In summary, the merger model is as follows:

$$\begin{cases} \begin{cases} y_{j-1} = S_{j-1} \\ y'_{j-1} = S'_{j-1} \end{cases} & \text{if } R_j \geq S_{j-1} + S'_{j-1}, \\ \begin{cases} y_{j-1} = \text{mid}\{S_{j-1}, R_j - S'_{j-1}, p_{j-1}R_j\} \\ y'_{j-1} = \text{mid}\{S'_{j-1}, R_j - S_{j-1}, p'_{j-1}R_j\} \end{cases} & \text{if } R_j < S_{j-1} + S'_{j-1}, \end{cases} \quad (9.3)$$

9.5.4 Divergence Scenario

A queuing model is also needed for a divergence where one stream of traffic splits into two. The divergence consists of one upstream link $j - 1$ and two downstream links (e.g., a mainline link j and an off-ramp j') (see Figure 9.7). Assume that during interval (t_{i-1}, t_i) link $j - 1$ has S_{j-1} vehicles to send, link j is able to receive R_j vehicles, and the off-ramp j' can receive R'_j vehicles. In addition, the turning movements are predetermined: β (e.g., 80%) traffic from link $j - 1$ goes to link j and β' (e.g., 20%) link j' , where $0 \leq \beta \leq 1$, $0 \leq \beta' \leq 1$, and $\beta + \beta' = 1$. Further assume that vehicles depart following a first in-first out queuing discipline, and if a vehicle fails to depart, it holds up all vehicles behind it. The question here is to determine the actual outflow of link $j - 1$, y_{j-1} , among which how many vehicles are destined for link j , y_j and how many are destined for the off-ramp, y'_j .

With these assumptions, the divergence queuing model is quite simple. First, the following relationships must hold:

$$\begin{cases} y_{j-1} = y_j + y'_j \leq S_{j-1}, \\ y_j = \beta y_{j-1} \leq R_j, \\ y'_j = \beta' y_{j-1} \leq R'_j. \end{cases} \quad (9.4)$$

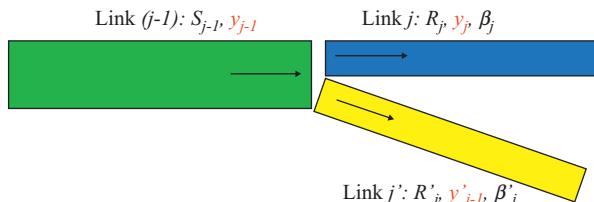


Figure 9.7 A freeway divergence.

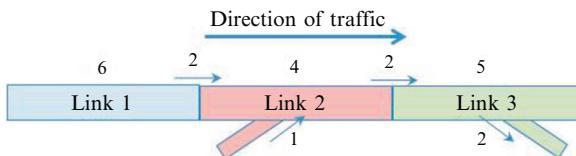
Hence,

$$\gamma_{j-1} = \min\{S_{j-1}, \frac{R_j}{\beta}, \frac{R'_j}{\beta'}\}.$$

Consequently, one obtains $\gamma_j = \beta \gamma_{j-1}$ and $\gamma'_j = \beta' \gamma_{j-1}$.

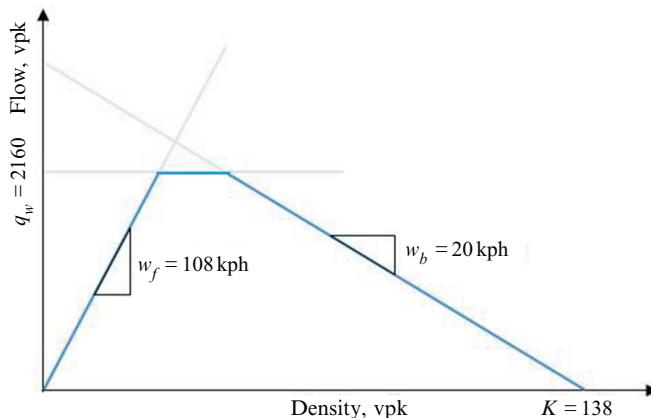
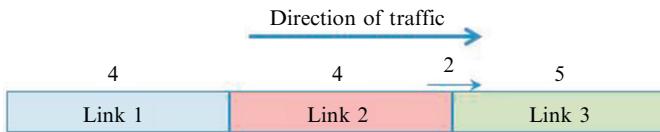
PROBLEMS

1. A small roadway system is illustrated in the figure below. The system consists of three links, each of which is 200 m long. In addition, there is an on-ramp at link 2 and an off-ramp at link 3. Currently, the storage in each link is indicated above the link. In the next time step (step size of 5 s), vehicles moving on are indicated at the borders, with arrows indicating where they go. Use the FLEFLO model to answer the following questions involved in a one-step simulation:

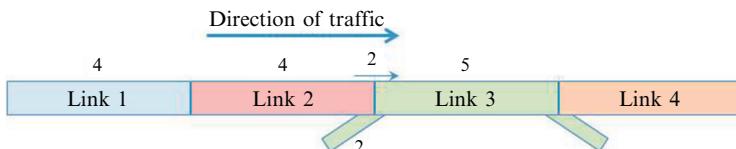


- a. What is the storage in link 2 at the end of the next step?
 - b. What is the density in link 2 at the end of the next step?
 - c. What is the equilibrium speed corresponding to this density?
2. Assume that the above road system and conditions remain the same. Use the KRONOS model to answer the following questions involved in a one-step simulation:
- a. What is the storage in link 2 at the end of the next step?
 - b. What is the density in link 2 at the end of the next step?
 - c. What is the equilibrium speed corresponding to this density if there is a Greenshields speed-density relationship with free-flow speed $v_f = 96$ km/h and jam density $k_j = 120$ vehicles per kilometer?
 - d. On the basis of your answer to (c), what is the corresponding equilibrium flow?
3. Assume that the road system and conditions are given in the figure below. The system consists of three links, each of which is 150 m long. Currently, the storage in each link is indicated above the link. In the next time step (step size of 5 s), two vehicles will move from link 2 to

link 3. Moreover, the triangular flow-density relationship illustrated in the figure below applies to each mainline link. Use the CTM to answer the following questions involved in a one-step simulation:



- What is the storage in link 2 at the end of the next step?
 - What is the density in link 2 at the end of the next step?
 - What is the equilibrium speed corresponding to this density?
 - On the basis of your answer to (c), what is the equilibrium flow?
4. Now more details of the road system become known. In addition to the conditions given above, the system includes a fourth link (link 4), an on-ramp at the upstream end of link 3, and an off-ramp at the downstream end of link 3. The on-ramp has two vehicles waiting to enter the mainline. Assume that the priority of the mainline versus the on-ramp is 2:1. At the off-ramp, 25% of vehicles in link 3 plan to exit.



- a. How many on-ramp vehicles may enter link 3 if it is able to receive five vehicles in the next step?
- b. How would your answer change if link 3 is able to receive three vehicles in the next step?
- c. If link 3 has room to accept only two vehicles, how many vehicles can actually enter link 3 from the mainline and the on-ramp?
- d. If the condition in (b) is true, how many vehicles want to enter link 4 in the next step?

CHAPTER 10

Simplified Theory of Kinematic Waves

The previous chapters have focused on analyzing the dynamic change of traffic states over time and space using the theory of waves. In particular, the method of characteristics applied to the LWR problem resulted in the identification of shock waves as a means to solve the continuity equation (i.e., the conservation law). In addition, numerical methods were discussed to approximate the solution.

Alternatively, Derivation V in Chapter 5 seems to suggest that the conservation law is self-guaranteed if a three-dimensional representation of traffic flow is used—that is,

$$q_x + k_t = \frac{\partial N^2(t, x)}{\partial x \partial t} - \frac{\partial N^2(t, x)}{\partial x \partial t} = 0.$$

The significance of the above equation is that there is no need to solve partial differential equations and find shock waves in order to analyze traffic dynamics. Instead, one only needs to count cars over time and space. As such, traffic dynamics is contained in these cumulative counts and can be extracted as the need arises.

This idea has been explored by Gordon F. Newell, who creatively integrated D/D/1 queuing theory into this idea to allow prediction from boundary and initial conditions. The result is known as the simplified theory of kinematic waves, published in a trio of papers [30–32] in the early 1990s. The first paper addresses the general theory, the second paper focuses on queuing at a freeway bottleneck, and the third paper deals with multidestination flows. Below I present the main points of the first and second papers interpreted from my own perspective. Interested readers are encouraged to use this chapter as a key to unlock the original papers for an enriched learning experience.

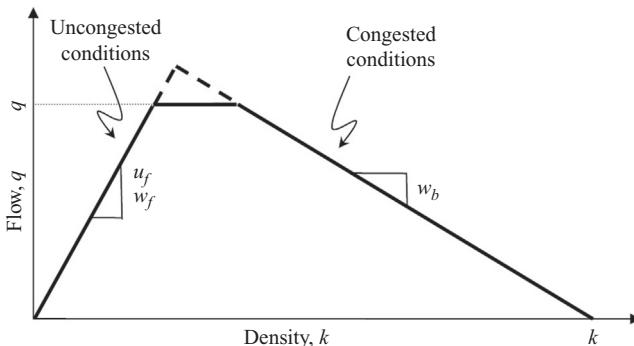


Figure 10.1 Triangular flow-density relationship.

10.1 TRIANGULAR FLOW-DENSITY RELATIONSHIP

The kinematic waves model was proposed as a (graphical) solution to the LWR model under a special condition: the underlying flow-density relationship is a triangular one with jam density K and capacity Q (see Figure 10.1).

From Figure 8.5, a point on the flow-density curve uniquely defines the operating condition of a stream of traffic. The speed of a kinematic wave carried by the traffic, w , is the tangent to the curve at this point. If the underlying flow-density relationship is triangular, finding kinematic wave speeds is greatly simplified. Actually, there are only two kinematic wave speeds: a forward wave speed w_f for all uncongested conditions (the left branch of the triangle) and a backward wave speed w_b for all congested conditions (the right branch). In addition, w_f happens to be the same as the free-flow speed v_f . As a special property of the triangular flow-density relationship, v_f applies to all uncongested conditions.

10.2 FORWARD WAVE PROPAGATION

Unlike conventional numerical models such as FREFLO, KRONOS, and the cell transmission model which keep track of cell storages $n(t_i, x_j)$ or equivalently cell densities $k(t_i, x_j)$, simplified kinematic waves model just counts vehicles at some predetermined locations. The outcome of the model is a set of cumulative flows representing the number of vehicles counted at these locations over time, $N(t, x_j), j \in (1, 2, \dots, J)$. These cumulative flows contain all the information that is needed to determine traffic dynamics over time and space.

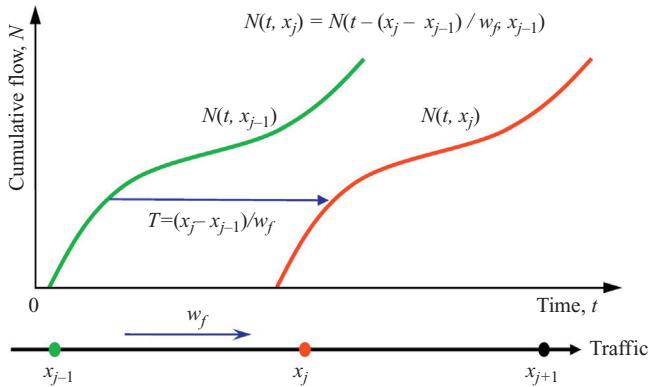


Figure 10.2 Forward wave propagation.

Suppose the cumulative flow recorded at location x_{j-1} over time t is $N(t, x_{j-1})$ and there is no congestion between x_{j-1} and x_{j+1} . The traffic will be dictated by (uncongested) upstream arrival from x_{j-1} , and these vehicles will arrive at downstream location x_j after a duration of $T = \frac{x_j - x_{j-1}}{v_f}$ if the vehicles preserve their order (i.e., first in, first-out). The traffic also carries a kinematic wave whose speed w_f happens to be v_f , as noted above, so it is equivalent to saying that the kinematic wave will propagate forward and arrive at x_j after $T = \frac{x_j - x_{j-1}}{w_f}$. Graphically, this forward wave propagation can be constructed as in Figure 10.2, where the profile $N(t, x_j)$ is simply a horizontal translation of profile $N(t, x_{j-1})$ to the right by T :

$$N(t, x_j) = N\left(t - T, x_{j-1}\right) = N\left(t - \frac{x_j - x_{j-1}}{w_f}, x_{j-1}\right).$$

10.3 BACKWARD WAVE PROPAGATION

Suppose the cumulative flow recorded at location x_{j+1} over time t is $N(t, x_{j+1})$ and there is congestion between x_{j-1} and x_{j+1} (see Figure 10.3). Then the kinematic wave carried by the traffic will propagate backward at speed w_b . Hence, the traffic condition at location x_j ($x_{j-1} < x_j < x_{j+1}$) will be dictated by downstream congestion. Consequently, cumulative flow at x_j , $N(t, x_j)$, will be a horizontal translation of $N(t, x_{j+1})$ to the right by $T = \frac{x_{j+1} - x_j}{w_b}$ shifted upward by a jam storage $n = K_j(x_{j+1} - x_j)$:

$$N(t, x_j) = N\left(t - T, x_{j+1}\right) + n = N\left(t - \frac{x_{j+1} - x_j}{w_b}, x_{j+1}\right) + K_j(x_{j+1} - x_j).$$

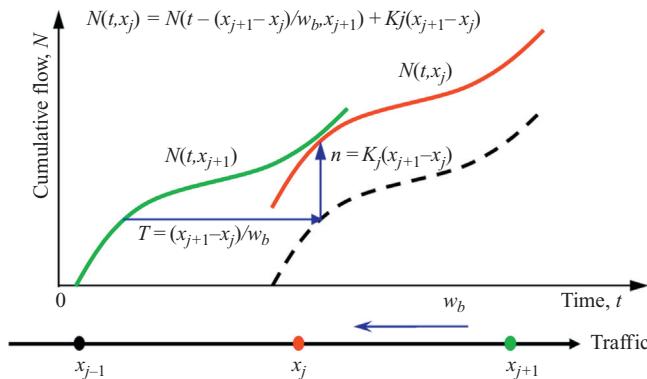


Figure 10.3 Backward wave propagation.

10.4 LOCAL CAPACITY

Suppose the cumulative flow to pass location x_j is $N(t, x_j)$ and the local capacity is Q_j . Since vehicles cannot be discharged beyond the capacity, this is equivalent to saying that the tangent to the profile $N(t, x_j)$ at any point should not exceed Q_j . Hence, the cumulative flow constrained by local capacity Q_j , $N^Q(t, x_j)$ is constructed as follows. Draw a line with slope Q_j from the right toward the profile $N(t, x_j)$ till the line is tangent to the profile. Any portion of the profile above the line is replaced by the latter. Continue the above process until no portion of the profile has a tangent greater than Q_j (see Figure 10.4).

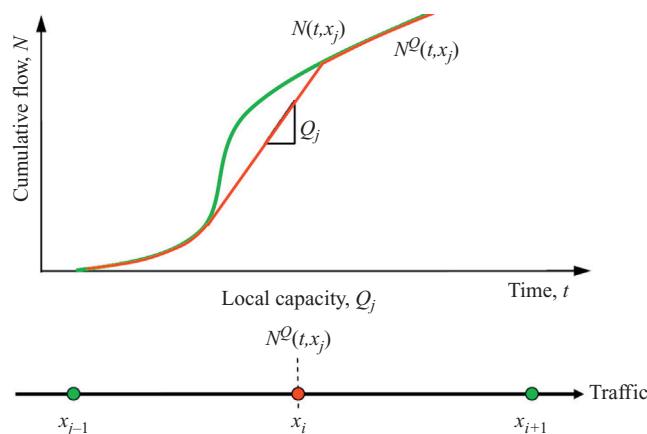


Figure 10.4 Flow constrained by local capacity.

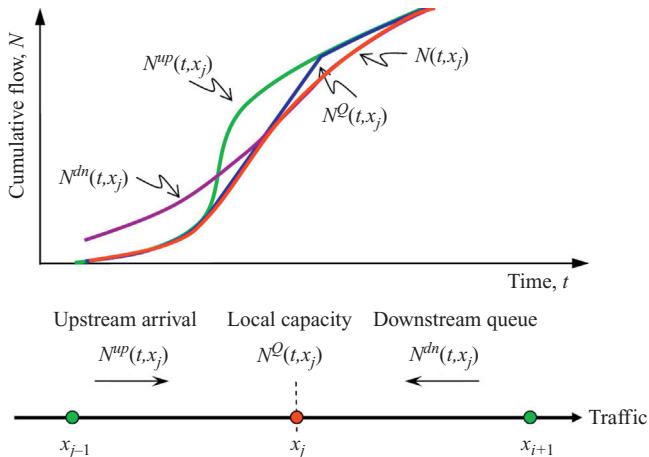


Figure 10.5 The minimum principle.

10.5 MINIMUM PRINCIPLE

Intuitively, the minimum principle means that any point on a roadway x_j cannot admit more vehicles than arrive from an upstream location $N^{up}(t, x_j)$, which is allowed by local capacity $N^Q(t, x_j)$, and which the downstream location is able to receive $N^{dn}(t, x_j)$. Graphically, this involves superimposing the above three curves on a single graph, and the cumulative flow that actually passes x_j , $N(t, x_j)$ is the lower envelope of the three (see Figure 10.5):

$$N(t, x_j) = \min\{N^{up}(t, x_j), N^Q(t, x_j), N^{dn}(t, x_j)\}.$$

10.6 SINGLE BOTTLENECK

In Figure 10.5, if there is an on-ramp at x_j , the location slightly downstream (to the right of x_j), x_j^+ , may be a bottleneck since both traffic streams from the upstream mainline and the on-ramp meet here. To keep track of arrival and departure flows, cumulative flow $N(t, x)$ will be replaced by two notations:

- cumulative arrival flow $A(t, x)$, which denotes cumulative flow having arrived at location x by time t waiting to pass x , and
- cumulative departure flow $D(t, x)$, which denotes cumulative flow having departed location x by time t .

Note that their difference $D(t, x) - A(t, x)$ gives the length of the queue at time t .

Central to the single bottleneck idea is to determine its cumulative arrival and departure flows, $A(t, x_j)$ and $D(t, x_j)$, given by

- upstream departure at earlier times $D(t, x_{j-1})$;
- downstream departure at earlier times $D(t, x_{j+1})$;
- on-ramp cumulative inflow $A_j(t)$;

From wave forward propagation (Figure 10.2), the cumulative flow arriving at location slightly upstream of the bottleneck (to the left of x_j), x_j^- is

$$A(t, x_j^-) = A\left(t - \frac{x_j - x_{j-1}}{w_f}, x_{j-1}\right).$$

Given on-ramp traffic $A_j(t)$, the cumulative flow arriving to the right of x_j is

$$N^{\text{up}}(t, x_j^+) = A(t, x_j^+) = A(t, x_j^-) + A_j(t).$$

From wave backward propagation (Figure 10.3), the cumulative flow allowed to depart is

$$N^{\text{dn}}(t, x_j^+) = N\left(t - \frac{x_{j+1} - x_j}{w_b}, x_{j+1}^-\right) + K_j(x_{j+1} - x_j).$$

Considering local capacity (Figure 10.4), the cumulative flow departing x_j^+ should not exceed $N_Q(t, x_j^+)$.

Therefore, on the basis of the minimum principle (Figure 10.5), the cumulative flow that actually departed at x_j^+ is

$$D(t, x_j^+) = \min\{N^{\text{up}}(t, x_j^+), N_Q(t, x_j^+), N^{\text{dn}}(t, x_j^+)\}.$$

If on-ramp traffic, $A_j(t)$, has priority over mainline traffic and can always bypass any queue at the bottleneck (this assumption is a limitation of the theory of kinematic waves since it eliminates queuing on ramps), then the cumulative departure flow to the left of x_j can be determined as

$$D(t, x_j^-) = D(t, x_j^+) - A_j(t).$$

The above procedure is illustrated in Figure 10.6.

10.7 COMPUTATIONAL ALGORITHM

With the above knowledge, traffic flow on a freeway involving multiple segments and bottlenecks can be numerically modeled as follows. First, the time and space are partitioned using the discretization scheme in

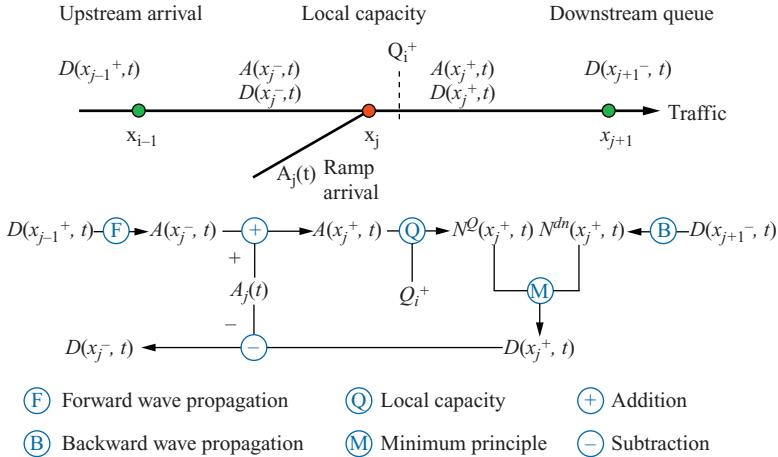


Figure 10.6 Single bottleneck.

Chapter 9, resulting the lattice shown in Figure 10.7. Next, starting from the initial conditions, one applies the numerical solution procedure outlined in Chapter 9. At each lattice point (t_i, x_j) , the cumulative arrival and departure flows are determined as follows:

1. Determine upstream arrival to x_j^- :

$$A(t_i, x_j^-) = D(t_i - \frac{x_j - x_{j-1}}{w_f}, x_{j-1}^+).$$

2. Determine upstream arrival to x_j^+ :

$$N^{up}(t_i, x_j^+) = A(t_i, x_j^+) = A(t_i, x_j^-) + A_j(t_i).$$

3. Apply capacity constraint at x_j^+ :

$$N^Q(t_i, x_j^+) = D(t_{i-1}, x_j^+) + Q_j^+ \times \Delta t,$$

where Q_j^+ is the capacity at x_j^+ and Δt is $t_i - t_{i-1}$.

4. Determine departure allowed by x_{j+1}^- :

$$N^{dn}(t_i, x_j^+) = N(t_i - \frac{x_{j+1}^- - x_j}{w_b}, x_{j+1}^-) + K_j(x_{j+1}^- - x_j).$$

5. Determine actual departure at x_j^+ :

$$D(t_i, x_j^+) = \min\{N^{up}(t_i, x_j^+), N^Q(t_i, x_j^+), N^{dn}(t_i, x_j^+)\}.$$

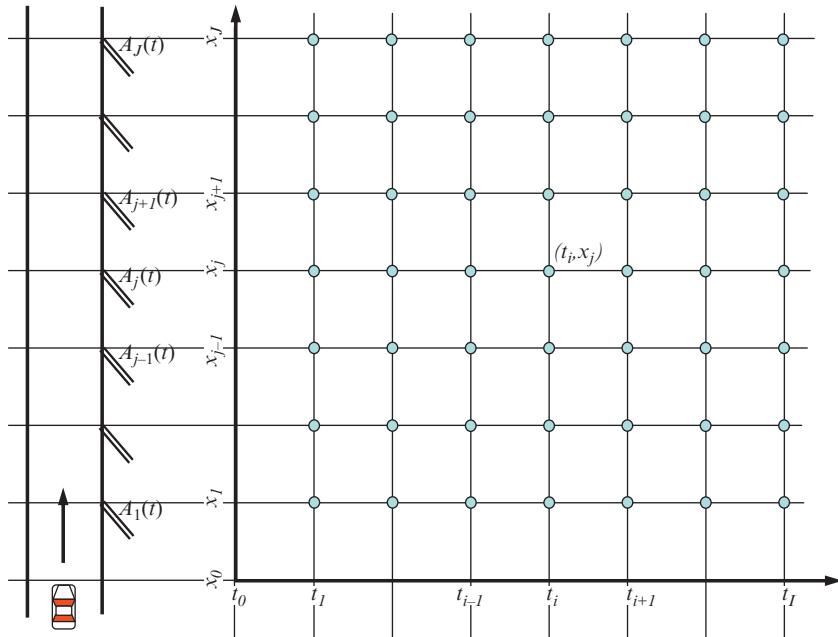


Figure 10.7 Lattice of kinematic waves.

6. Determine actual departure at x_j^- :

$$D(t_i, x_j^-) = D(t_i, x_j^+) - A_j(t_i)$$

7. Proceed to the next lattice point (t_i, x_{j+1}) .

Repeat the above steps at lattice point (t_i, x_{j+1}) till the end (t_i, x_J) . Then advance time to t_{i+1} and start over again from (t_{i+1}, x_1) to (t_{i+1}, x_J) . Repeat the above steps till all the lattice points have been traversed.

The result of this computational algorithm is a set of cumulative arrival and departure flows:

$$A(t_1, x_1^-), D(t_1, x_1^-), A(t_1, x_1^+), D(t_1, x_1^+)$$

...

$$A(t_1, x_J^-), D(t_1, x_J^-), A(t_1, x_J^+), D(t_1, x_J^+)$$

$$A(t_2, x_1^-), D(t_2, x_1^-), A(t_2, x_1^+), D(t_2, x_1^+)$$

...

$$A(t_I, x_J^-), D(t_I, x_J^-), A(t_I, x_J^+), D(t_I, x_J^+).$$

10.8 FURTHER NOTE ON THE THEORY OF KINEMATIC WAVES

The above discussion summarizes the first two papers of Newell's simplified theory of kinematic waves [30, 31] involving bottlenecks with on-ramps only. The third paper [32] takes off-ramps into consideration, and hence multiple destination flows. Discussion of this subject is quite involved, and readers are encouraged to read the original paper for full information. In addition, supplementary information on the simplified theory of kinematic waves can be found in Son [33] and Hurdle and Son [34] for model validation and extraction of information of traffic dynamics and in Ni [35] and Ni et al. [36] for extension of the theory and associated computational algorithms.

Though Newell's theory involves partitioning a highway into a series of segments, the lengths of these segments do not necessarily have to be equal and small. Such a partitioning is necessary only at locations where capacity changes (e.g., lane drop), there is an on-ramp, and there is an off-ramp. Therefore, the resulting number of segments can be much less than in cell-based models such as FREFLOW, KRONOS, and the cell transmission model, whose accuracy relies on cell size (length of segment). Consequently, the computation and storage requirements can be significantly reduced.

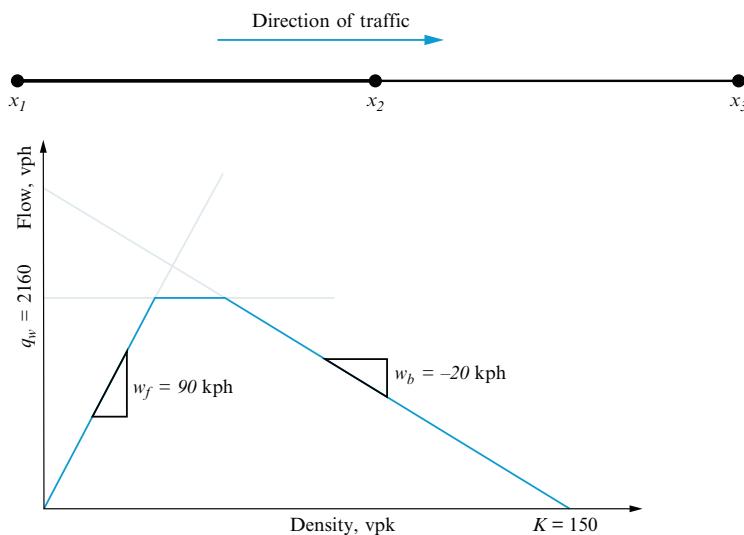
In addition to the assumption of a triangular flow-density relationship, another limitation of the theory of kinematic waves is its assumption that on-ramp traffic has priority over mainline traffic and can always bypass any queue at a bottleneck. Consequently, the theory of kinematic waves is unable to model network traffic where queuing at ramps has to be accounted for. A further attempt to address this issue can be found in Ni [35] and Ni et al. [36], where on-ramp and off-ramp queuing models were proposed, on the basis of which the theory of kinematic waves was extended to network flows.

PROBLEMS

1. A 1.8-km link AB connects nodes A and B. The free-flow speed is 30 m/s and a triangular flow-density relationship is assumed for this link. How long does it take for traffic passing node A to arrive at node B if there is no congestion in this link?
2. Link AB above is followed by link BC of length 1 km. The same flow-density relationship applies to link BC with backward wave propagation speed -5 m/s . If an accident occurred at node C at 5:00 p.m., when will drivers at node B notice the impact of the accident?

3. A uniform freeway link x_1x_3 is 6 km long as illustrated in the figure below. The triangular flow-density relationship given below applies to this link. Node x_2 is the midpoint of this link. A segment of data is given below where D_1 , D_2 , and D_3 are cumulative traffic counts at x_1 , x_2 , and x_3 , respectively. On the basis of the simplified theory of kinematic waves, complete a one-step simulation by answering the following questions:

Time	D1	D2	D3
8:50	1965	1950	1551
8:51	1970	1957	1555
8:52	1978	1961	1560
8:53	1982	1968	1566
8:54	1987	1973	1571
8:55	1991	1980	1578
8:56	1996	1984	1583
8:57	2000	1990	1589
8:58	2008	1996	1594
8:59	2012	2000	1600
9:00	2018	2005	1605
9:01	2024		1610

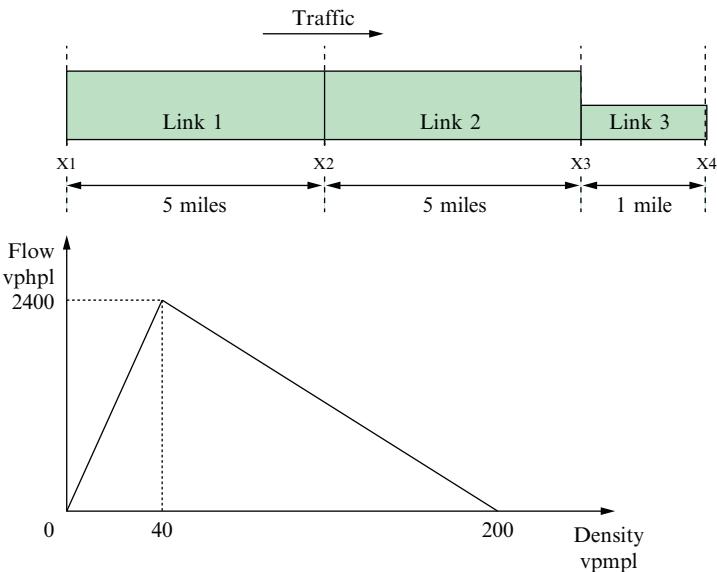


- a. How many vehicles are expected to arrive x_2 from 9:00 to 9:01?
- b. Dictated by capacity only, what is the cumulative number of vehicles that are allowed to pass x_2 by 9:01?
- c. What is the jam storage in x_2x_3 —that is, the number of vehicles that can be stored in x_2x_3 at jam density?

- d. On the basis of the condition in x_2x_3 only, how many vehicles are allowed to enter x_2x_3 by 9:01 at most?
- e. What is the cumulative number of vehicles that actually pass x_2 by 9:01?
4. A freeway corridor consists of three links whose physical properties are tabulated below. Also provided in the figure below is the underlying flow-density relationship of the freeway corridor. Assume the freeway corridor was initially empty and subsequent traffic arrival from the upstream end is as given in the second table below. Use the simplified theory of kinematic waves to simulate traffic evolution on this freeway corridor. You may use an Excel spreadsheet or a computer program such as MATLAB if necessary.

	Link 1	Link 2	Link 3
Lanes	2	2	1
Free-flow speed	60	60	60
Capacity	4800	4800	2400
Jam	400	400	200

Time	Flow (vehicles/h)
0:00:00	0
1:00:00	120
2:00:00	240
3:00:00	480
4:00:00	600
5:00:00	1200
6:00:00	1500
7:00:00	1800
8:00:00	3000
9:00:00	3600
10:00:00	1800
11:00:00	1200
12:00:00	1500
13:00:00	900
14:00:00	1200
15:00:00	1500
16:00:00	2400
17:00:00	3600
18:00:00	2100
19:00:00	1500
20:00:00	1200
21:00:00	600
22:00:00	240
23:00:00	0



CHAPTER 11

High-Order Models

The macroscopic traffic flow models discussed so far, including both analytical and numerical models, have been focused on the LWR model [24, 25] and its variants. At the center of these models is mass or vehicle conservation, which can be mathematically expressed as a first-order partial differential equation:

$$\frac{\partial k(x, t)}{\partial t} + \frac{\partial q(x, t)}{\partial x} = 0,$$

where k and q are density and flow, which depend on time t and space x . Hence, these models are referred to as *first-order* models.

Common to first-order models is their prediction of a shock wave when two kinematic waves meet. Consequently, a vehicle crossing the shock wave has to change its speed abruptly, which is physically impossible. This limitation, together with other undesirable features, has led many researchers to seek more realistic models to represent traffic dynamics. Naturally, these efforts gave rise to *high-order* dynamic traffic flow models.

11.1 HIGH-ORDER MODELS

In essence, the conservation law takes several forms, among which mass or vehicle conservation is perhaps the simplest. Other forms of the law are conservation of linear momentum and conservation of energy, which involve high-order partial differential equations. If a model involves such equations, it is classified as a high-order model, a few examples of which are described below.

11.1.1 PW Model (1971)

Proposed by Payne [37] and independently by Whitham [38], the PW model consists of a system of two equations: the first is the conservation of mass given in the LWR model, and the second equation is derived from the Navier-Stokes equation of motion for a one-dimensional compressible flow with a pressure and a relaxation term.

$$\begin{cases} \frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0, \\ \frac{\partial v}{\partial t} = -v \frac{\partial v}{\partial x} - \lambda(v - V_e(k)) - \frac{1}{k} \frac{dP}{dk} \frac{\partial k}{\partial x}, \end{cases}$$

where v is traffic speed, $V_e(k)$ is the equilibrium speed-density relationship, $P(k)$ is traffic pressure, and λ is a coefficient. Note that FREFLO presented in Chapter 9 is a numerical solution to the PW model.

11.1.2 Phillips's Model (1979)

On the basis of kinetic theory, Phillips [39] developed a model which incorporates mass conservation, momentum conservation, and energy conservation:

$$\begin{cases} \frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0, \\ \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = \lambda(V_e(k) - v) - \frac{1}{k} \frac{\partial P}{\partial x}, \\ \frac{\partial P}{\partial t} + u \frac{\partial P}{\partial x} = \lambda[k(V_e(k) - v)^2 + (P_e - P)] - 3P \frac{\partial v}{\partial x}, \end{cases}$$

where P_e is the equilibrium traffic pressure, and everything else is as defined above.

11.1.3 Kühne's Model (1984)

Kühne [40, 41] also proposed a model by considering sound speed and viscosity:

$$\begin{cases} \frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0, \\ \frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} = \lambda(V_e(k) - v) - \frac{c_0^2}{k} \frac{\partial k}{\partial x} + \eta \frac{\partial^2 v}{\partial x^2}, \end{cases}$$

where c_0 is sound speed and η is a viscosity constant.

11.1.4 Kerner and Konhäuser's Model (1993)

Kerner and Konhäuser [42] showed that given an initially homogeneous traffic flow, regions of high density and low average speed (clusters of cars) can spontaneously appear. These high-density regions can move either with or against the flow of traffic, and two clusters with different speeds, widths, and amplitudes merge when they meet, resulting in a single cluster. The continuum flow model adopted is in the following form:

$$\begin{cases} \frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0, \\ \frac{\partial v}{\partial t} = \lambda(V_e(k) - v) - \frac{c_0^2}{k} \frac{\partial k}{\partial x} + \frac{1}{k} \frac{\eta \partial v}{\partial x}. \end{cases}$$

11.1.5 Model of Michalopoulos et al. (1993)

Michalopoulos et al. [43] proposed a model which does not require the use of an equilibrium speed-density relationship. Traffic friction at interrupted flows and changing geometries is also addressed through the use of a viscosity term. Tests with field data and comparison with existing models suggested that the proposed model is more accurate and computationally more efficient than existing models.

$$\begin{cases} \frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0, \\ \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = \frac{1}{\tau}(\nu_f - v) - G \frac{\partial v}{\partial t} - \nu k^\beta \frac{\partial k}{\partial x}, \end{cases}$$

where $G = \mu k^\varepsilon g$, μ , ν , ε , and β are all constant parameters, and ν_f is the free-flow speed.

11.1.6 Zhang's Model (1998)

Zhang [44] proposed a nonequilibrium traffic flow model which is based on both empirical evidence of traffic flow behavior and basic assumptions about drivers' reactions to stimuli. By assuming an equilibrium speed-density relationship and introducing a disturbance propagation speed, the model includes the LWR model as a special case and removes some of its deficiencies. Unlike existing high-order continuum models, this model eliminates "wrong-way travel" because in this model traffic disturbances are always propagated against the traffic stream.

$$\begin{cases} \frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0, \\ \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = \lambda(V_e(k) - v) - k(V'_e(k))^2 \frac{\partial k}{\partial x}. \end{cases}$$

11.1.7 Model of Treiber et al. (1999)

Treiber et al. [45] derived macroscopic traffic equations from specific gas-kinetic equations, and the resulting partial differential equations for vehicle density and average speed contain a nonlocal interaction term which is very favorable for a fast and robust numerical integration, so several thousand

freeway kilometers can be simulated in real time.

$$\begin{cases} \frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0, \\ \frac{\partial \nu}{\partial t} + \nu \frac{\partial \nu}{\partial x} = \lambda(V_e(k) - \nu) + \frac{1}{k} \frac{\partial k A \nu^2}{\partial x} - \frac{V_e A(k)}{\tau A(k_j)} \left[\frac{k_\alpha T \nu}{1 - \frac{k_\alpha}{k_j}} \right]^2 B(\delta_\nu), \end{cases}$$

where $A = A(k)$ is a density-dependent function, k_α is the density at point x_α ahead of x , $B(\delta_\nu)$ is a macroscopic interaction term, and $V_e(k)$ is the normal equilibrium speed-density relationship.

11.2 RELATING CONTINUUM FLOW MODELS

Starting from mass or vehicle conservation, a variety of continuum flow models have been developed by the inclusion of additional assumptions. Generally, these models can be summarized by the following model [46]:

$$\begin{cases} \frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = g(t, x), \\ \frac{\partial \nu}{\partial t} + \nu \frac{\partial \nu}{\partial x} = \frac{1}{\tau}(V_e(k) - \nu) + \frac{1}{k} \frac{\partial P}{\partial x}, \end{cases}$$

where $V_e(k, \nu)$ is the generalized equilibrium speed-density relationship, $P(k, \nu)$ is the traffic pressure, and τ is the relaxation time, which is the time constant of regulating the traffic speed ν to the equilibrium speed V_e . $g(t, x)$ is the net ramp inflow. For a highway without on-ramps or off-ramps, $g(t, x) = 0$.

Each of the above-mentioned continuum flow models can be viewed as a special case of the general model when different traffic pressure P , relaxation time τ , and generalized equilibrium speed U_e are applied. For example,

- the LWR model results if $\tau = 0$ and $P = 0$;
- the PW model results if $P = -\frac{V_e(k)}{2\tau}$ with $V_e(k, \nu) = V_e(k)$;
- Phillips's model results if $P = k\Theta$ with $\Theta = \Theta_0(1 - \frac{k}{k_j})$, where k_j is the jam density;
- Kerner and Konhäuser's model results if $P = k\Theta_0 - \eta \frac{\partial \nu}{\partial x}$;
- the model of Michalopoulos et al results if $P = \frac{\nu}{\beta+2} k^{\beta+2}$, where ν is an anticipation parameter, β is a dimensionless constant, and $V_e(k) = \nu_f$, where ν_f is the free-flow speed;
- Zhang's model resulted if $P = \frac{1}{3} k^3 V_e'^2(k)$, where $V_e'(k) = \frac{dV_e(k)}{dk}$;

- the model of Treiber et al. results if $P = Akv^2$, where $A = A(k)$ is a density-dependent function, and $V_e(k, v) = V_e(k)\{1 - \frac{A}{A(k_j)}[\frac{k_\alpha T v}{1 - \frac{k_\alpha}{k_j}}]^2 B(\delta_v)\}$.

11.3 RELATIVE MERITS OF CONTINUUM MODELS

Daganzo [47] noted that, as a first-order continuum flow model, the LWR model is proposed for dense traffic with an equilibrium and it is flawed for light traffic. This is because, when passing is allowed, the LWR model fails to recognize that the preferred speed for each vehicle varies over time and the desired speeds among a group of vehicles vary as well. These variations can cause a platoon to disperse in a way that is not predicted by the LWR model. When passing is allowed, the LWR model produces unsatisfactory results in the following three aspects. First, the LWR model predicts an abrupt speed change when a vehicle passes through a shock wave, an action that is unrealistic in the real world. Second, the LWR model fails to predict instabilities of stop-start traffic. Third, the LWR model assumes zero reaction time, which does not happen in the real world. Readers are referred to Daganzo's original paper for full information.

Given these deficiencies, the continuum flow models developed so far have been trying to fix the deficiencies, and almost all of these models follow the direction of incorporating a momentum conservation equation. An early attempt to fix the deficiencies in the LWR model was made by Prigogine [48], who proposed a kinetic model incorporating a speed distribution to address platoon dispersion. A decade later, Payne [37] and Whitham [38] proposed a dynamic model, the so-called PM model, trying to smooth out the discontinuity in speed change across shock waves. A momentum equation was introduced in this model to describe the structure of a shock wave. This seminal work has inspired many thoughts regarding analytical explanation of shock wave behavior, and thus has spawned several variants, among which are those of Phillips [39], Kühne [40, 41], Kerner and Konhäuser [42], Michalopoulos et al. [43], Zhang [44], and Treiber et al. [45].

Several deficiencies are found in the PW model [47]. First, it does not remove all the shock waves. Second, as reported by del Castillo et al. [49], vehicles in the PW model can adjust their speeds in response to disturbance from behind, while in reality vehicles typically respond to their leaders. Third, the PW model incorporates a momentum equation

which is derived from a car-following model. This momentum equation does not consider second-order and higher-order terms of spacings and speeds, which may not be negligible when spacings and speeds are not slowly varying. Fourth, the PM model as well as other high-order models always produces wave speeds that are greater than traffic speeds. This is an unattractive property of macroscopic models because it implies that future conditions of a vehicle are partially decided by what happens behind it. Fifth, the strength that high-order models smooth out shocks turns out to be these models' weakness. This is because any model that attempts to smooth all the discontinuities must sometimes predict negative speeds and such negative speeds observed in computer models cannot be removed by convergent numerical approximation methods. Sixth, but probably not the last, high-order models involve more complex partial differential equations and more variables, which increases computational complexity, and are more difficult to calibrate and implement. Given these limitations, many researchers [47, 50, 51] tend to believe that high-order models, despite their added complexity and additional parameters, might not be superior to the LWR model.

11.4 TAXONOMY OF MACROSCOPIC MODELS

[Figure 11.1](#) shows a rather simple and incomplete taxonomy which relates macroscopic traffic flow models to each other. The figure starts with the basic principle, conservation laws, which takes the forms of mass conservation, momentum conservation, and energy conservation.

Mass conservation and a functional flow-density relationship (typically derived from an equilibrium speed-density relation) constitute the core of the LWR model. This model is classified as a first-order model since it involves a first-order partial differential equation. Numerical models derived from LWR models are indicated as double-line boxes in the left panel. These models include KRONOS, the kinematic waves model (though this is a graphical solution involving discrete space but continuous time), and the cell transmission model. The kinematics waves model was further extended to network traffic by Ni [35] and Ni et al. [36].

Central to high-order models are equations of mass and momentum conservation. These models include the models of Payne [37] and Whitham [38], Phillips [39], Kühne [40, 41], Kerner and Konhäuser[42], Michalopoulos [43], Zhang [44], Treiber et al. [45], etc. FREFLO is a numerical model derived from the model of Payne [37] and Whitham [38].

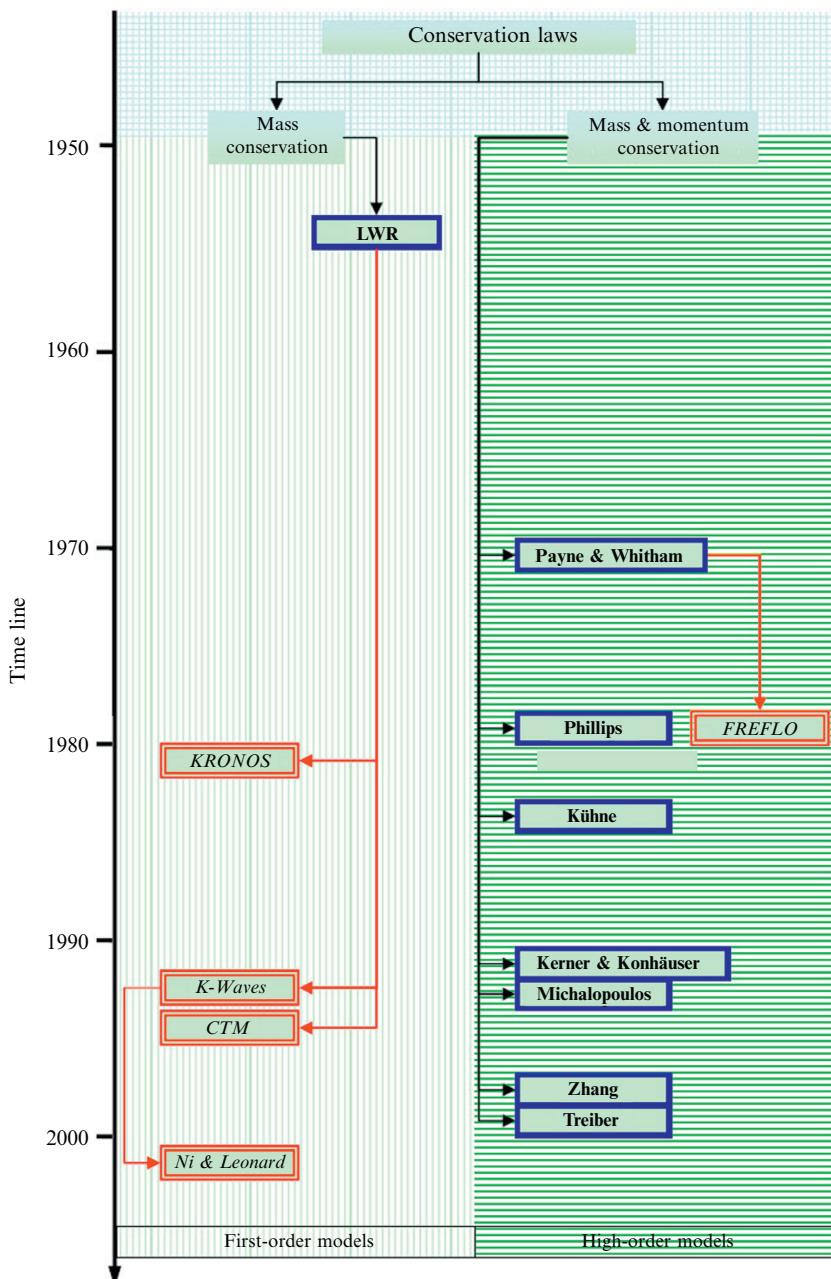


Figure 11.1 Taxonomy of macroscopic models. CTM, cell transmission model; K-waves, kinematic waves.

PROBLEMS

1. Name a few similarities and differences between first-order and high-order models.
2. What are the major reasons that motivated the exploration of high-order models?
3. Highlight a few drawbacks of high-order models.
4. Compare the equation of momentum conservation in the PW model and Phillips's model.
 - a. Comment on how they differ.
 - b. Which one is more general?
 - c. Can one of them be derived from the other?
5. The high-order models introduced in this book include an equation representing the conservation of vehicles. This equation implicitly assumes a uniform freeway segment without on-ramps and off-ramps—that is, traffic flows in and out via mainline lanes but not ramps. How would the equation of vehicle conservation change if traffic from and to ramps is considered?

CHAPTER 12

Microscopic Modeling

The models presented in Chapters 4–11 emphasize collective and average behavior of vehicles (e.g., flow, speed, and density), and consider traffic flow as a compressible fluid. Central to these models are the relationships among flow, speed, and density as well as how they vary dynamically over time and space. Such models are termed *macroscopic*, and they are capable of capturing the amount of “fluid” (i.e., number of vehicles) flowing into and out of roadway segments over time, rather than tracking each and every vehicle as it moves along the roadway.

In contrast, *microscopic* models emphasize the behavior of individual vehicles, and are capable of capturing the motion of and interaction among these vehicles. Unlike macroscopic models, which treat vehicles as a fluid, microscopic models represent a driver-vehicle unit as a particle without mass. Such a particle is sometimes referred to as an “active” particle since it is capable of making decisions based on rules stipulated in microscopic models.

12.1 MODELING SCOPE AND TIME FRAME

Depending on the geographical scope and time frame involved, driving decisions can be categorized at three levels—namely, strategic, tactical, and operational. Driving decisions at the *strategic* level involve a large geographical scope and a long time frame. For example, Figure 12.1 illustrates the decision-making scenario faced by a driver who is about to travel from the University of Massachusetts Amherst (point C) to Boston (point D). The driver has at least three options:

1. Interstate 90 (Massachusetts Turnpike). The bottom route, which is the fastest route if there is no congestion, and the toll is about \$5.
2. Route 2. The top route, which is a scenic, rural highway that is rarely congested.
3. Route 9. The middle route, which is the shortest route, but it goes through many town centers and traffic signals.

This scenario constitutes a **route-choice** decision that involves a geographical scope of about 100 km and a time frame of a few hours. A

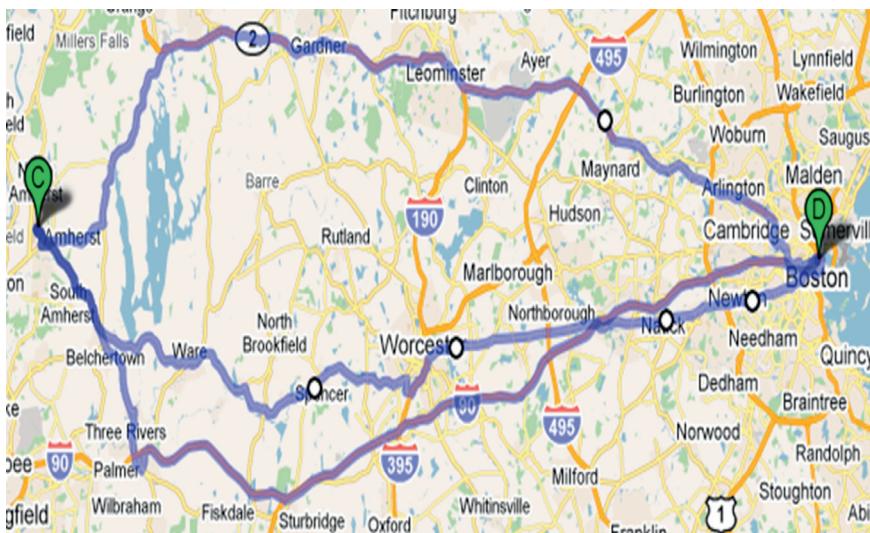


Figure 12.1 Making a decision at the strategic level.

microscopic model that describes how drivers make a route choice decision is called a route-choice model. Such a model is typically a discrete choice model which chooses one of a set of options based on some utilities and constraints.

After the driver has chosen a route (e.g., Massachusetts Turnpike) and is traveling down the road, a *tactical* decision will have to be made sooner or later that involves a medium geographical scope and a medium time frame. For example, [Figure 12.2](#) illustrates that the driver needs to decide when and where to change to the side lane in preparation for using the upcoming exit. Such a case constitutes a **lane-changing** decision with a geographical scope of a few kilometers and a time frame of a few minutes. Again, a lane-changing model is typically a discrete choice model that determines the choice of a target lane from available options based on the driver's objective and constraints.

An *operational* decision involves the driver's operational control of the vehicle in order to ensure safety and maintain mobility within a small geographical scope and a short time frame. For example, [Figure 12.3](#) illustrates that the driver (in the circled vehicle) is following another vehicle in a context of a geographical scope of tens of meters and a time frame of a few seconds. The driver needs to make a **car-following** decision on how

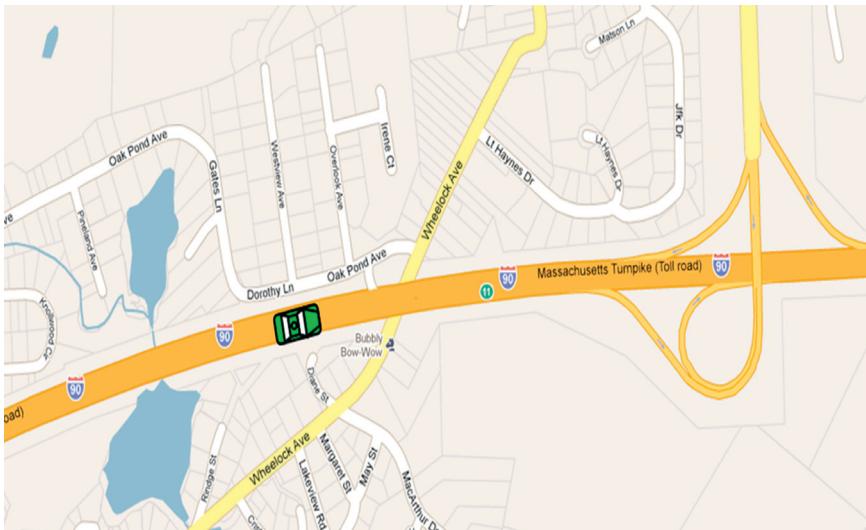


Figure 12.2 Making a decision at the tactical level.



Figure 12.3 Making a decision at the operational level.

to operate his or her vehicle (e.g., determine speed and acceleration in the next second) so as to avoid colliding with the leading vehicle. Meanwhile, if the driver feels stressed following the slow leading vehicle, the driver may want to change to another lane to improve his or her mobility. As such, the driver makes a **gap-acceptance** decision by looking for gaps in the adjacent lane and switching to that lane when an acceptable gap becomes available.

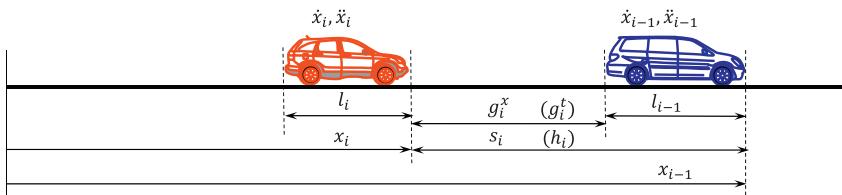


Figure 12.4 A car-following scenario.

Therefore, on the basis of the geographical scope and time frame involved, microscopic models can fall into one of the following three broad categories:

1. at the strategic level: route-choice models;
2. at the tactical level: lane-changing models;
3. at the operational level: car-following and gap-acceptance models.

12.2 NOTATION

The chapters that follow will emphasize car-following models. More specifically, drivers' operational control when following another vehicle on a single-lane highway will be considered where no passing is allowed. Before the formal discussion of car-following models, it is helpful to summarize the notation to be used. [Figure 12.4](#) illustrates two vehicles traveling on a one-lane highway. These vehicles ($1, 2, \dots, i - 1, i, i + 1, \dots, I$) are numbered cumulatively with lower-numbered vehicles in front—for example, vehicle 1 leads vehicle 2. The locations or displacements of vehicles are measured from a common but arbitrary reference point.

i	vehicle ID, $i = 1, 2, \dots, I$.
$x_i(t)$	the location of vehicle i at time t .
$\dot{x}_i(t)$	the speed of vehicle i at time t .
v_i	desirable speed that driver i is willing to travel at whenever possible.
$\ddot{x}_i(t)$	the acceleration of vehicle i at time t .
A_i	the maximum acceleration that vehicle i is able to apply. $A_i > 0$.
B_i	the maximum deceleration that vehicle i is able to apply. $B_i < 0$.
l_i	the length of vehicle i .
τ_i	the perception-reaction time of driver i .
$s_i(t)$	the spacing between vehicle i and the leading vehicle $i - 1$ at time t .
$g_i^x(t)$	the distance between vehicle i and the vehicle in front of it at time t .
$h_i(t)$	the headway between vehicle i and the vehicle in front of it at time t .
$g_i^t(t)$	the time gap between vehicle i and the vehicle in front of it at time t .

12.3 BENCHMARKING SCENARIOS

The upcoming chapters will introduce some microscopic car-following models. These models were formulated with a variety of modeling philosophies and appeared in different forms. It would be very interesting and informative if these models could be cross-compared on the basis of a common ground. Such a process is called benchmarking, two scenarios of which are set up here, one being microscopic and the other macroscopic.

12.3.1 Microscopic Benchmarking

Microscopic benchmarking employs a concrete example consisting of a set of hypothetical driving regimes. The purpose of microscopic benchmarking is to illustrate the performance of these car-following models in different regimes so that their operational control under various conditions can be examined.

The example involves two vehicles: a leading vehicle $i-1$ and a following vehicle i . The motion of the leader is predetermined and that of the follower is governed by the car-following model under study. Initially ($t = 0$), vehicle $i-1$ stands still at 5000 m from the reference point ($x_{i-1}(0) = 5000$ m, $\dot{x}_{i-1}(0) = 0$ m/s, and $\ddot{x}_{i-1}(0) = 0$ m/s²). Vehicle i , which is also still ($\dot{x}_i(0) = 0$ m/s and $\ddot{x}_i(0) = 0$ m/s²), stands somewhere near the reference point, with the exact location to be determined case by case in different car-following models. When the scenario starts ($t > 0$), vehicle $i-1$ remains still, while vehicle i starts to move. Since vehicle $i-1$ is far ahead, vehicle i is entitled to accelerate freely to satisfy its driver's desire for mobility. At time $t = 100$ s, vehicle i is at somewhere about $x_i(100) \approx 2770$ m. At this moment, a third vehicle previously moving in the adjacent lane at 24 m/s changes to the subject lane at location 2810 m and takes over as the new leading vehicle, assuming ID $i-1$ —that is, $x_{i-1}(100) = 2810$ m, $\dot{x}_{i-1}(100) = 24$ m/s, and $\ddot{x}_{i-1}(100) = 0$ m/s². This change is designed to mimic the effect that a vehicle cuts in in front of another vehicle with a spacing of about 40 m. Meanwhile, the previous, stationary leading vehicle is removed from the road. The new leading vehicle keeps moving at that speed up to $t = 200$ s, and then undergoes deceleration at a rate of $\ddot{x}_{i-1} = -3$ m/s² until it comes to a complete stop. After that, vehicle $i-1$ remains stopped up to $t = 300$ s. Then, it begins to accelerate at a constant rate of $\ddot{x}_{i-1} = 2$ m/s², and eventually settles at its full speed of $\dot{x}_{i-1} = 36$ m/s. At time $t = 400$ s, the vehicle starts to decelerate again at

a constant rate of $\ddot{x}_{i-1} = -3 \text{ m/s}^2$ until it comes to another full stop, and remains there. During all the time, the motion of the follower i is completely stipulated by the car-following model. The above scenario is formulated as follows.

$$\begin{cases} x_{i-1} = 5000 \text{ m}, \dot{x}_{i-1} = 0 \text{ m/s}, \ddot{x}_{i-1} = 0 \text{ m/s}^2 & \text{when } 0 \text{ s} \leq t < 100 \text{ s}, \\ x_{i-1} = 2810 \text{ m}, \dot{x}_{i-1} = 24 \text{ m/s} & \text{when } t = 100 \text{ s}, \\ \ddot{x}_{i-1} = 0 \text{ m/s}^2 & \text{when } 100 \text{ s} \leq t < 200 \text{ s}, \\ \ddot{x}_{i-1} = -3 \text{ m/s}^2 & \text{when } 200 \text{ s} \leq t < 208 \text{ s}, \\ \ddot{x}_{i-1} = 0 \text{ m/s}^2 & \text{when } 208 \text{ s} \leq t < 300 \text{ s}, \\ \ddot{x}_{i-1} = 2 \text{ m/s}^2 & \text{when } 300 \text{ s} \leq t < 318 \text{ s}, \\ \ddot{x}_{i-1} = 0 \text{ m/s}^2 & \text{when } 318 \text{ s} \leq t < 400 \text{ s}, \\ \ddot{x}_{i-1} = -3 \text{ m/s}^2 & \text{when } 400 \text{ s} \leq t < 412 \text{ s}, \\ \ddot{x}_{i-1} = 0 \text{ m/s}^2 & \text{when } t \geq 412 \text{ s}. \end{cases}$$

The driving regimes involved in the above process include the following:

- Start-up: Vehicle i starts to move from standstill, when the process begins ($t > 0$ s).
- Speedup: After start-up, vehicle i continues to accelerate to higher speeds ($0 \text{ s} < t < 100 \text{ s}$).
- Free flow: As vehicle i speeds up, it settles at its desired speed if it is unimpeded ($0 \text{ s} < t < 100 \text{ s}$).
- Cutoff: A sudden decrease in spacing owing to the new leader $i - 1$ cutting in ($t = 100$ s).
- Following: Vehicle i has to adopt vehicle $i - 1$'s speed so as to avoid a collision ($100 \text{ s} < t < 200 \text{ s}$).
- Stop and go: Vehicle i is forced to stop and go owing to vehicle $i - 1$'s brief stopping ($200 \text{ s} \geq t \leq 300 \text{ s}$).
- Trailing: Vehicle i is following a speeding leader ($300 \text{ s} < t < 400 \text{ s}$).
- Approaching: Vehicle i is getting close to a slower or stationary leader ($400 \text{ s} \geq t < 420 \text{ s}$). item Stopping: Vehicle i tries to stop behind a stationary object separated by a minimum spacing ($t \geq 420 \text{ s}$).

This scenario involves a series of tests in a single driving process. Rather than seeking “the best” model, our focus here is to analyze whether a model makes physical sense by facing these tests. Therefore, the reality check includes the following items:

- Start-up: Whether the model itself is sufficient to start the vehicle up without involving any additional, external logic.

- Speedup: Whether the model generates speed and acceleration profiles that make physical sense.
- Free flow: Whether the model settles at its desired speed without overshooting or undershooting.
- Cutoff: Whether the model loses control or, if not, responds with a reasonable control maneuver.
- Following: Whether the model is able to adopt the leader's speed and follow the leader at a reasonable distance.
- Stop and go: Whether the model is able to stop the vehicle safely behind its leader and start moving again when the leader resumes motion.
- Trailing: Whether the model is able to speed up normally without being tempted to speed up by its speeding leader—that is, a vehicle is attracted to excessively high speeds by its speeding leader.
- Approaching: Whether the model is able to adjust the vehicle properly when the intervehicle spacing closes up.
- Stopping: Whether the model is able to stop the vehicle properly behind a stationary object separated by a minimum spacing, without overshooting or undershooting, and causing speed and acceleration to return to zero naturally when stopped, etc.

Note that the starting position of the follower i is determined by trial and error such that the vehicle moves to $x_i \approx 2770$ m at $t = 100$ s, at which point the vehicle should have reached its desired speed $v_i = 30$ m/s. The sudden appearance of the new leader $i - 1$ at $x_{i-1} = 2810$ m leaves a spacing of about 40 m between the two vehicles, which is a little more than the distance traversed during one perception-reaction time. Drivers would normally back up a little in this situation and then identify a comfortable spacing to start car following.

12.3.2 Macroscopic Benchmarking

Macroscopic benchmarking employs a set of empirical data obtained from Georgia 400, a toll road with freeway by design located in Atlanta, Georgia, USA. The data contain 1-years' worth of field observations at one station across four lanes. The fundamental diagram (i.e., mathematical and/or graphical presentation that illustrates the collective behavior of traffic flow) observed at this station is depicted in [Figure 12.5](#). This figure contains a set of four plots that illustrate speed-density, speed-flow, flow-density, and speed-spacing relationships. The “cloud” contains field observations of flow, speed, and density aggregated to 5 min intervals. To highlight the average

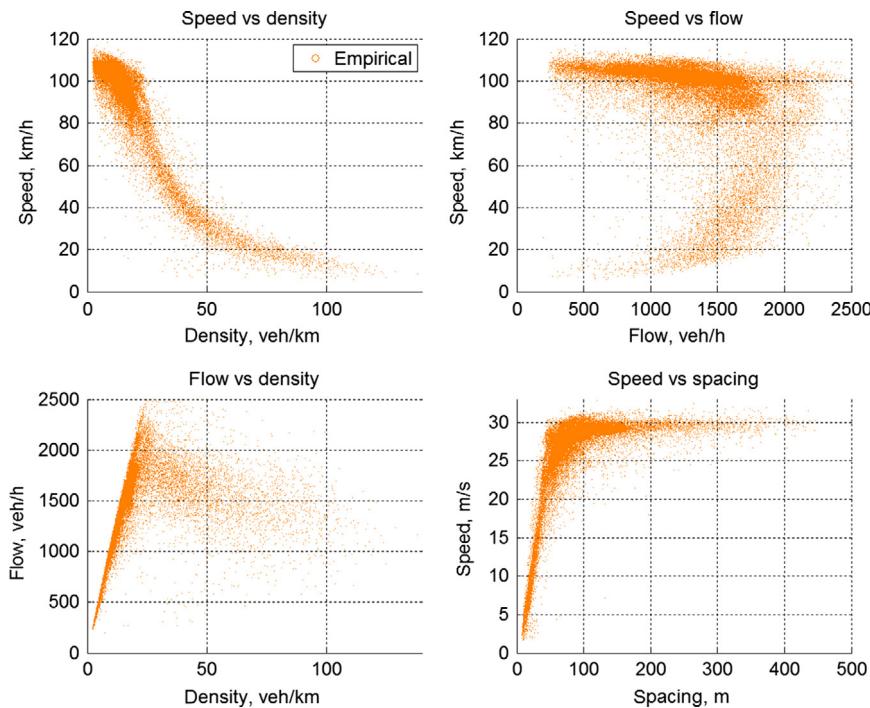


Figure 12.5 Empirical fundamental diagram observed from the field.

behavior of traffic flow, the observations in the cloud are further aggregated with respect to density, and the result is shown as circles.

PROBLEMS

1. What are the major differences between microscopic modeling and macroscopic modeling?
2. Route choice, lane change, and car following are all about decision making. Elaborate the difference among these decision making processes.
3. Identify the following traffic flow characteristics with use of the empirical fundamental diagram illustrated in this book:
 - a. Free-flow speed
 - b. Capacity condition (capacity, optimal speed, and optimal density)
 - c. Jam density

CHAPTER 13

Pipes and Forbes Models

As the beginning discussion on car-following models, this chapter introduces two simple models—that is, the Pipes model and the Forbes model, both of which are derived from drivers' daily driving experiences.

13.1 PIPES MODEL

The Pipes model [52] is based on a safe driving rule coined in the California Vehicle Code:

A good rule for following another vehicle at a safe distance is to allow yourself at least the length of a car between your vehicle and the vehicle ahead of you for every ten mile per hour of speed at which you are traveling.

Referring to Figure 13.1 and putting the safety rule in mathematical language, we get

$$g_i^x(t)_{\min} = [(x_{i-1}(t) - x_i(t)) - l_{i-1}]_{\min} = (s_i(t) - l_{i-1})_{\min} = \frac{\dot{x}_i(t)}{0.447 \times 10} l_i, \quad (13.1)$$

where $\dot{x}_i(t)$ is in meters per second (1 mile per hour is approximately 0.447 m/s), and $g_i^x(t)$, $x_{i-1}(t)$, and $x_i(t)$ are measured in meters. The Pipes model is formulated as

$$s_i(t)_{\min} = \frac{l_i}{4.47} \dot{x}_i(t) + l_{i-1}. \quad (13.2)$$

If we assume a vehicle length of 6 m, the model reduces to

$$s_i(t)_{\min} = 1.34 \dot{x}_i(t) + 6 \quad (13.3)$$

or

$$h_i(t)_{\min} = 1.34 + \frac{6}{\dot{x}_i(t)}. \quad (13.4)$$

13.1.1 Applications of the Pipes Model

The Pipes model can be applied in many ways, the two foremost of which are automatic driving and computer simulation.

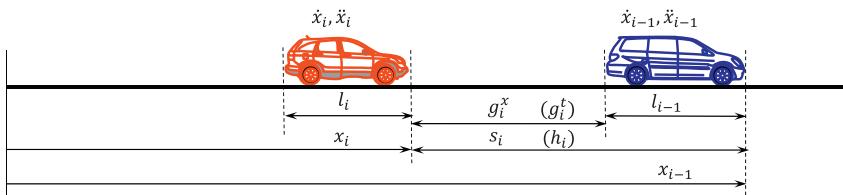


Figure 13.1 A car-following scenario.

Automatic Driving

Perhaps the simplest form of automatic driving is cruise control. As an in-vehicle system, cruise control automatically controls the speed of a motor vehicle (by taking over the control of the throttle) so that the vehicle maintains a constant speed set by its driver. Cruise control makes it easier to drive on long road trips, and hence is a popular car feature. As more and more vehicles join the traffic and the road becomes crowded, the driver has to switch cruise control on and off so frequently that cruise control becomes less useful. To adapt to the dynamics of the vehicle in front, it is desirable that the cruise control system be able to adjust speed accordingly (rather than cruising at a preset speed) to maintain a safe car-following distance. Hence, an adaptive or autonomous cruise control system has been developed. With the aid of distance sensors such as radar or laser sensors, autonomous cruise control allows the vehicle to slow down when approaching another vehicle and accelerate to the preset speed when traffic conditions permit. To make this happen, the system requires an internal logic which relates the vehicle speed to the distance to the vehicle in front. Simple car-following models such as the Pipes model can be employed as the basis of such an internal logic. More specifically, Equation 13.1 can be rearranged as follows:

$$\dot{x}_i(t) \leq \frac{0.447 \times 10}{l_i} g_i^x(t). \quad (13.5)$$

For a vehicle length of 6 m, the above control logic becomes $\dot{x}_i(t) \leq 0.745 g_i^x(t)$. Therefore, the autonomous cruise control works as follows. At any moment t , the distance sensor measures the gap between the two vehicles $g_i^x(t)$. Then, the target speed that the vehicle needs to adapt to is set as $0.745 g_i^x(t)$ or less.

Obviously, the target speed can easily go out of bound as the gap $g_i^x(t)$ becomes sufficiently large. Therefore, it is necessary to set an upper bound

to the target speed, which is usually referred to as the desirable speed v_i . Therefore, the target speed is actually the minimum of (1) the desirable speed v_i and (2) the speed constrained by the vehicle in front $\dot{x}_i(t)$:

$$\dot{x}_i(t) \leq \min\{v_i, \frac{0.447 \times 10}{l_i} g_i^x(t)\}. \quad (13.6)$$

Computer Simulation

The Pipes model can also be used to simulate a platoon of vehicles moving on a one-lane highway. Before the simulation starts, the following variables need to be initialized—that is, a value needs to be assigned to each of them:

- l_i length of vehicle $i \in \{1, 2, \dots, I\}$
- τ_i perception-reaction time of driver i
- v_i desired speed of driver i
- ΔA_i maximum acceleration of vehicle i
- ΔB_i maximum deceleration of vehicle i
- Δt simulation time step

At time step j , the displacement x and speed v of each vehicle are updated:

```

FOR i = 1:I
    s(j,i) = x(j-1,i-1) - x(j-1,i);
    s_min(j,i) = l(i) * (v(j-1,i)/(0.447 * 10) + 1);
    IF s(j,i) < s_min(j,i)
        v(j,i) = MAX([0, v(j-1,i) - dB_i]);
    ELSE
        v(j,i) = MIN([v_i, v(j-1,i) + dA_i]);
    END
    x(j,i) = x(j-1,i) + v(j,i) * dt;
END

```

In the above code segment, the actual spacing between vehicle i and its leading vehicle, $s(j, i)$, is computed as the difference of their locations in the previous time step. The minimum safe spacing, $s_{\min}(j, i)$ is determined according to the California Vehicle Code. Then, $s(j, i)$ is compared against $s_{\min}(j, i)$. If $s(j, i)$ is less than $s_{\min}(j, i)$, one should reduce the speed of the vehicle by ΔB_i , but should not go beyond 0. Otherwise, one should increase the speed of the vehicle by ΔA_i without exceeding its desired speed v_i .

Then, one should update the position of the vehicle, advance time by one step, and continue with the next vehicle.

Note that car-following models used for automatic control and computer simulation have different objectives. The objective of automatic control is to guarantee safety but achieve mobility (e.g., arrive at the destination without delay). As such, automatic control calls for “an ideal (or the best) driver/model” that is able to operate the vehicle in the best way. In contrast, the purpose of computer simulation is to reproduce part of the real world as realistically as possible. Consequently, computer simulation necessities “a representative driver/model” that is able to mimic the behavior of day-to-day driving, which is usually not perfect.

13.1.2 Properties of the Pipes Model

In mathematical modeling, it is always interesting to understand how a system’s microscopic behavior relates to its macroscopic behavior, or alternatively to interpret the microscopic basis of a macroscopic phenomenon. In traffic flow theory, microscopic car-following models are typically related to macroscopic speed-density relationships and further the fundamental diagram.

Typically, the linkage between microscopic and macroscopic models can be addressed in two ways. One approach is to run a simulation based on the microscopic model. Such a microscopic simulation typically involves random variables such as perception-reaction time, desired speed, and acceleration rate. As a result, simulation results vary in different runs. Hence, the macroscopic behavior implied by the microscopic model can be obtained by a statistical analysis of these simulation results.

The other approach is analytical—that is, one tries to aggregate or integrate the microscopic model (which typically involves ordinary differential equations) under some equilibrium or steady-state assumptions. If a system is in the steady state, any property of the system is unchanging in time. More specifically, a traffic system in the steady state would consist of homogeneous vehicles which exhibit uniform behavior over time and space. Therefore, under steady-state conditions, vehicles lose their identities (e.g., $\tau_i \rightarrow \tau$ and $l_i \rightarrow l$), vehicles travel at uniform speed (i.e., $\dot{x}_i = \dot{x}_j \rightarrow v$ and $\ddot{x}_i \rightarrow 0$), drivers’ desired speeds converge to the free-flow speed (i.e., $v_i \rightarrow v_f$), and the vehicle spacing $s_i(t)$ reduces to s , which, in turn, is replaced by the reciprocal of traffic density $\frac{1}{k}$. Uniform vehicle length l is equivalent to the reciprocal of jam density k_j —that is, $\frac{1}{k_j}$. Hence, the Pipes model reduces to

$$\frac{k_j}{k} = \frac{\nu}{4.47} + 1 \text{ or } \nu = 4.47 \left(\frac{k_j}{k} - 1 \right), \quad (13.7)$$

where k is measured in vehicles per meter and ν is measured in meters per second. With $q = k \times \nu$, the above speed-density relationship gives rise to the following flow-density and speed-flow relationships:

$$q = 4.47(k_j - k) \quad (13.8)$$

and

$$\nu = \frac{q}{k_j - 0.22q}. \quad (13.9)$$

Equations 13.7–13.9 constitute the mathematical representation of the fundamental diagram implied by the Pipes model.

13.2 FORBES MODEL

Rather than ensuring a safe distance between vehicles as the Pipes model does, Forbes [53, 54] stipulates that

To ensure safety, the time gap between a vehicle and the vehicle in front of it should be always greater than or equal to reaction time.

This safety rule can be formulated as

$$g_i^t(t) = h_i(t) - \frac{l_i}{\dot{x}_i} \geq \tau_i. \quad (13.10)$$

For a reaction time of 1.5 s and a vehicle length 6 m, the model becomes

$$h_i(t) \geq 1.5 + \frac{6}{\dot{x}_i} \quad (13.11)$$

or

$$s_i(t) \geq 1.5\dot{x}_i + 6. \quad (13.12)$$

This is very similar to the Pipes model except for a slight difference in the coefficient of the speed term, which is interpreted as perception-reaction time τ_i . Therefore, the Pipes model and the Forbes model are essentially equivalent and can be generically expressed as

$$s_i(t) \geq \tau_i\dot{x}_i + l_i, \quad (13.13)$$

where τ_i and vehicle length l_i are model parameters. Note that applications and properties of the Pipes model discussed above apply to the Forbes model. In addition, the fundamental diagram implied by the Pipes and Forbes models can be generically expressed as

$$\nu = \frac{1}{\tau k} - \frac{l}{\tau}, \quad (13.14)$$

$$q = \frac{1}{\tau} - \frac{l}{\tau} k, \quad (13.15)$$

$$\nu = \frac{ql}{1 - \tau q}, \quad (13.16)$$

where τ is the average perception-reaction time and l is the average vehicle length.

13.3 BENCHMARKING

Since the Pipes and Forbes models are essentially equivalent, the following discussion addresses only the Pipes model with the understanding that the result applies to the Forbes model as well. Microscopic benchmarking refers to the scenario presented in Section 12.3.1 and macroscopic benchmarking refers to the scenario presented in Section 12.3.2.

13.3.1 Microscopic Benchmarking

For convenience, the Pipes model is rearranged as

$$\dot{x}_i(t + \Delta t) = \frac{s_i(t) - l_i}{\alpha}, \quad (13.17)$$

where Δt is the simulation time step and α is a constant resulting from unit conversion ($\alpha = 1.34$ if speed is in meters per second and $l_i = 6$ m).

First, the model has a problem with vehicle acceleration. We refer to the microscopic benchmarking scenario presented in Section 12.3.1, and suppose that initially the leading vehicle is located at $x_{i-1}(0) = 5000$ m and the subject vehicle is at $x_i(0) = -102$ m and both vehicles are standing still. When the simulation begins, vehicle i starts to move according to the Pipes model. A spacing of $s_i(0) = 5102$ m results in a speed of about 3800 m/s at the next time step (assuming $\Delta t = 1$ s), which requires an acceleration of 3800 m/s². It follows that an infinite speed and acceleration would result if there is no leading vehicle in front. Therefore, the following external

logic has to be imposed on the Pipes model in order to limit its maximum acceleration:

$$\ddot{x}_i(t) = \frac{\dot{x}_i(t + \Delta t) - \dot{x}_i(t)}{\Delta t} \leq A_i, \quad (13.18)$$

where A_i is the maximum acceleration of vehicle i —for example, $A_i = 4 \text{ m/s}^2$. With this addition, the Pipes model loses its mathematical elegance which favors a one-equation-for-all formulation. Even though an external logic is added, the Pipes model still has a problem with the maximum speed. For example, it is true that the acceleration no longer exceeds A_i , but the vehicle can still reach unrealistically high speeds—for example, $\dot{x}_i = 196 \text{ m/s}$ when $s_i = 590 \text{ m}$. Therefore, another external logic has to be imposed to limit the speed:

$$\dot{x}_i \leq v_i, \quad (13.19)$$

where v_i is driver i 's desired speed. The third problem is unrealistic deceleration. For example, at time $t = 424$, vehicle i is located at about $x_i = 8734 \text{ m}$ moving at speed $\dot{x}_i = 30 \text{ m/s}$, while vehicle $i - 1$ stops at $x_{i-1} = 8762 \text{ m}$. According to the Pipes model, vehicle i 's speed at the next step would be $\dot{x}_i \approx 16.42 \text{ m/s}$. As such, the deceleration rate is $\ddot{x}_i = -13.58 \text{ m/s}^2$. Hence, a third external logic has to be imposed to limit maximum deceleration B_i (e.g., -6 m/s^2):

$$\ddot{x}_i(t) = \frac{\dot{x}_i(t + \Delta t) - \dot{x}_i(t)}{\Delta t} \geq B_i. \quad (13.20)$$

However, this addition introduces a new problem. For example, vehicle i 's speed at the next step becomes $\dot{x}_i = 30 - 6 = 24 \text{ m/s}^2$ and its location is $x_i = 8758 \text{ m}$. This would leave a spacing of $s_i = 4 \text{ m}$, which is less than a vehicle length $l_{i-1} = 6 \text{ m}$ —that is, vehicle i has collided with vehicle $i - 1$. Unfortunately, there is no easy remedy to the problem except for accepting the unrealistic deceleration behavior.

The benchmarking result of the Pipes model with the constraints in Equations 13.18–13.20 is plotted in Figure 13.2. The performance of the constrained Pipes model is summarized as follows, and the discussion is based on the benchmarking scenario:

- Start-up: the model is able to start the vehicle up from standstill. See Figure 13.2 when $t > 0 \text{ s}$.

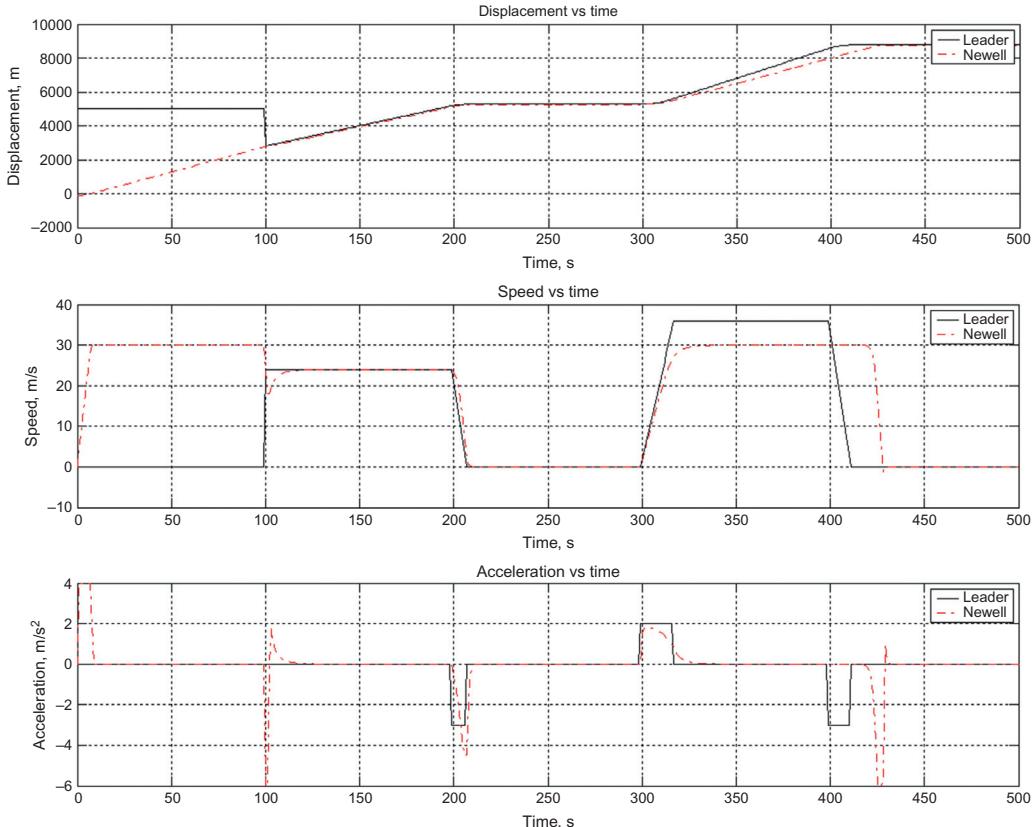


Figure 13.2 Microscopic benchmarking of the Pipes model.

Table 13.1 Microscopic benchmarking parameters of the Pipes model

l_i	v_i	τ_i	α	-
6 m	30 m/s	1.0 s	1.34	-
A_i	B_i	$x_i(0)$	$\dot{x}_i(0)$	$\ddot{x}_i(0)$
4.0 m/s ²	6.0 m/s ²	-120 m	0 m/s	0 m/s ²

- Speedup: the model is able to speed up the vehicle. However, its acceleration profile (i.e., acceleration as a function of speed) is unrealistic because the vehicle is able to retain maximum acceleration at high speeds. Normally, maximum acceleration is available only when a vehicle starts up. As the vehicle speeds up, acceleration decreases and eventually vanishes when the vehicle achieves its desired/cruising speed. See [Figure 13.2](#) when $0 \text{ s} < t < 100 \text{ s}$.
- Free flow: an external logic has to be imposed to limit the maximum speed under the free-flow condition. See [Figure 13.2](#) when $0 \text{ s} < t < 100 \text{ s}$.
- Cutoff: the model retains control and responds reasonably when a vehicle cuts in in front. See [Figure 13.2](#) around $t = 100 \text{ s}$.
- Following: the model is able to adopt the leader's speed and follow the leader by a reasonable distance. See [Figure 13.2](#) when $100 \text{ s} < t < 200 \text{ s}$.
- Stop and go: the model is able to stop the vehicle safely behind its leader and start it moving when the leader departs. See [Figure 13.2](#) when $200 \text{ s} \geq t \leq 300 \text{ s}$.
- Trailing: the model is able to speed up normally without being tempted to speed up by its speeding leader. See [Figure 13.2](#) when $300 \text{ s} < t < 400 \text{ s}$.
- Approaching: the model is unable to decelerate properly when approaching a stationary vehicle at a distance. The vehicle might collide with its leader when maximum deceleration is imposed. See [Figure 13.2](#) when $400 \text{ s} \geq t < 420 \text{ s}$.
- Stopping: this portion is invalid since approaching fails. See [Figure 13.2](#) when $t \geq 420 \text{ s}$.

The above benchmarking is based on the set of parameters in [Table 13.1](#), and the outcome may differ for a different set of parameters.

13.3.2 Macroscopic Benchmarking

The fundamental diagram implied by the Pipes model is plotted in [Figure 13.3](#) against empirical observations. The “cloud” contains 5 min observations of flow, speed, and density, the circles are empirical observa-

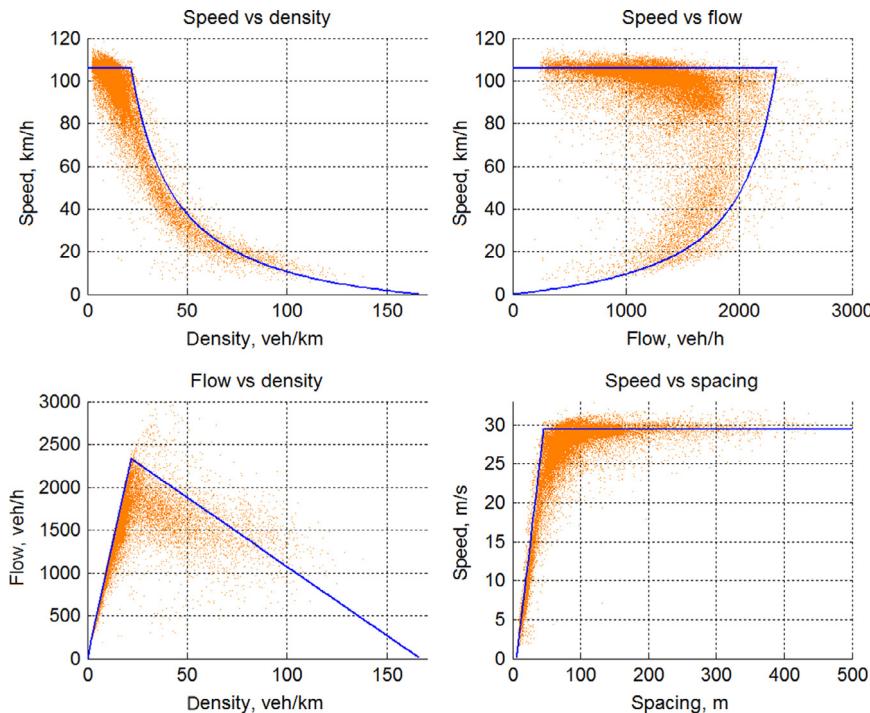


Figure 13.3 Fundamental diagram implied by the Pipes model.

tions aggregated with respect to density, and the curves are the equilibrium relationships implied by the Pipes model.

These curves roughly fit the empirical data in the middle to upper range of density (e.g., $k > 20$ vehicles per kilometer), but do not apply to the low-density range (e.g., $k < 20$ vehicles per kilometer). It appears that the Pipes model is designed to literally describe car-following behavior. Cases when the leading vehicle is absent have to be handled by an external logic. In addition, the Pipes model predicts that traffic speed would increase to infinity as density approaches zero.

The above benchmarking is based on the set of parameters in [Table 13.2](#), and the outcome may differ for a different set of parameters.

Table 13.2 Macroscopic
benchmarking parameters
of the Pipes model

$\alpha(\tau)$	I
1.34 s	6 m

Since the Forbes model is essentially the same as the Pipes model, the above benchmarking results apply to the Forbes model as well.

PROBLEMS

1. Prove that the Pipes and Forbes models are mathematically equivalent.
2. From the perspective of the spatial domain, the Pipes model suggests that drivers add a space gap of at least one car length for every 10 miles per hour of speed at which they travel. From the perspective of the temporal domain, the Forbes model requires that drivers leave a time gap of at least one perception-reaction time.
 - a. Since the two models are mathematically equivalent, what is the equivalent perception-reaction time that the Pipes model implies when the model is translated into the temporal domain? Assume a vehicle length of 6 m.
 - b. Assume a perception-reaction time of 1.5 s and a vehicle length of 6 m. Translate the Forbes model into the spatial domain and elaborate the equivalent driving rule—for example, drivers need to add a space gap of at least one car length for every x miles per hour of speed at which they travel.
3. An autonomous cruise control system is designed as follows. At any moment t , the onboard sensor measures the distance from this vehicle to the vehicle in front. Then the target speed of the vehicle is set as the minimum of (1) the distance multiplied by 0.8 and (2) the desired speed of 108 km/h. Assume all vehicles are controlled by this logic and the vehicle length is uniformly 7.5 m. What is the maximum number of vehicles that can pass a point of highway in 1 h?
4. Perform the following analysis based on the Forbes model under the assumption that the desired speeds of all drivers are uniformly 108 km/h, perception-reaction times are uniformly 1.5 s, and vehicle lengths are uniformly 5 m.
 - a. Find the capacity condition implied by the Forbes model (capacity, optimal speed, and optimal density).
 - b. If the uniform desired speed drops to 96 km/h, how would your answer change?
 - c. If the uniform perception-reaction time becomes 1 s but the uniform desired speed is held at 108 km/h, how would your answer change?
 - d. Illustrate and indicate the direction of change graphically on the basis of the underlying flow-density relationship.

5. Perform a one-step simulation based on the following conditions: Two cars are traveling in the same lane on a freeway. The length of both vehicles is $l_{i-1} = l_i = 6$ m. Lane change is not considered in this problem. At time t , the leading vehicle $i - 1$ is traveling at a speed of $\dot{x}_{i-1}(t) = 72$ km/h and the following vehicle i is traveling at a speed of $\dot{x}_i(t) = 108$ km/h. The spacing between the two vehicles (measured from front bumper to front bumper) is $s_i(t) = 40$ m. The perception-reaction time of the following driver is $\tau_i = 1.5$ s. Assume that the acceleration rate and deceleration rate are $\ddot{x}_i = 1 \text{ m/s}^2$.
- a. Use the Pipes model to predict the speed that the following driver will adopt after a perception-reaction time.
 - b. Use the Forbes model to predict the speed that the following driver will adopt after a perception-reaction time.

CHAPTER 14

General Motors Models

An anecdote has it that General Motors' CEO Charles Wilson once said "what's good for General Motors is good for America." It turned out that this statement was misquoted, and the true version dates back to 1953, when Wilson, who was appointed as the Secretary of Defense by President Eisenhower, was at his confirmation hearings before the Senate Armed Services Committee (source, Wikipedia):

During the hearings, when asked if as Secretary of Defense he could make a decision adverse to the interests of General Motors, Wilson answered affirmatively. But added that he could not conceive of such a situation "because for years I thought what was good for the country was good for General Motors and vice versa.

While what Wilson actually said in this anecdote is not of great interest here, what is important is the role that General Motors played in the history of traffic flow theory. Back in the 1950s, General Motors sponsored a team of scientists in its research laboratories, from which pioneering work was done that broke the ground for traffic flow theory. At the foremost of such efforts was the family of General Motors models (referred to as the GM models hereafter).

14.1 DEVELOPMENT OF GM MODELS

GM models [55, 56] assume that a driver's control maneuver is a result of not only external stimuli such as the dynamics of the subject vehicle and its leading vehicle, but also the driver's sensitivity. Hence such a relationship can be expressed as

$$\text{Response} = f(\text{sensitivity}, \text{stimuli}).$$

When formulating the above relationship, GM researchers chose the subject vehicle's acceleration (deceleration is negative acceleration) produced after a reaction time, $\ddot{x}_i(t + \tau)$, as the response (see Figure 14.1). The consideration of stimuli and sensitivity evolved over time and resulted in a family of models.

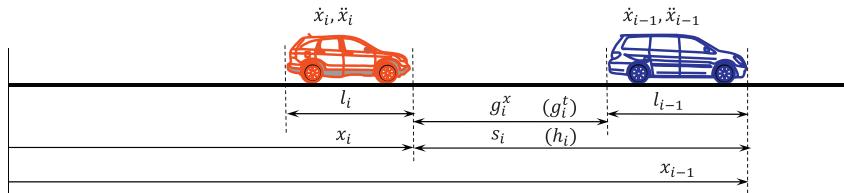


Figure 14.1 A car-following scenario.

14.1.1 GM1

Originally, General Motors researchers observed that drivers responded to the relative speed between the subject vehicle i and its leading vehicle $i - 1$, $\dot{x}_{i-1}(t) - \dot{x}_i(t)$. If sensitivity is treated as a coefficient that is multiplicative to the stimulus, the subject driver's operational control can be formulated as

$$\ddot{x}_i(t + \tau_i) = \alpha[\dot{x}_{i-1}(t) - \dot{x}_i(t)]. \quad (14.1)$$

The above model is the first-generation model which can be used to interpret some car-following phenomena effectively. For example, when the subject vehicle approaches its leading vehicle (e.g., $\dot{x}_i(t) = 120$ km/h and $\dot{x}_{i-1}(t) = 100$ km/h), the relative speed is negative, and hence the driver will decelerate since $\ddot{x}_i(t + \tau_i) < 0$ (assuming that α is a positive constant). In contrast, if the subject vehicle is falling behind its leading vehicle (e.g., $\dot{x}_i(t) = 100$ km/h and $\dot{x}_{i-1}(t) = 120$ km/h), the subject vehicle will accelerate because the relative speed now becomes positive. However, the model has difficulty distinguishing scenarios with large and small car-following distances. For example, the model predicts the same deceleration response to the following two scenarios:

- Scenario 1: $\dot{x}_i(t) = 120$ km/h, $\dot{x}_{i-1}(t) = 100$ km/h, and $s_i(t) = 50$ m
- Scenario 2: $\dot{x}_i(t) = 120$ km/h, $\dot{x}_{i-1}(t) = 100$ km/h, and $s_i(t) = 5000$ m

Since both scenarios have a speed difference of -20 km/h, intuitively, the subject driver in scenario 1 would brake much harder than the driver in scenario 2 because the former is facing an imminent collision.

14.1.2 GM2

The effect of spacing motivated General Motors researchers to choose different sensitivity coefficients, and hence the second-generation model resulted:

$$\ddot{x}_i(t + \tau_i) = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} [\dot{x}_{i-1}(t) - \dot{x}_i(t)]. \quad (14.2)$$

Field experiments revealed that the sensitivity coefficient α ranges between 0.17 and 0.74. In GM2, a high sensitivity value α_1 is chosen when the two vehicles are close, while a low sensitivity value α_2 is employed when the two vehicles are far apart.

14.1.3 GM3

The effect of spacing was partially addressed in GM2 because one has to frequently calibrate the sensitivity coefficient depending on car-following distances. The inconvenience seemed to suggest that spacing should be explicitly included in the model, which led to the formulation of the third-generation model:

$$\ddot{x}_i(t + \tau_i) = \alpha \frac{[\dot{x}_{i-1}(t) - \dot{x}_i(t)]}{[x_{i-1}(t) - x_i(t)]}. \quad (14.3)$$

Although the issue of spacing has been suppressed, another problem pops up. The model is unable to predict any difference between the following scenarios:

- Scenario 1: In downtown Amherst, one vehicle is following another at a spacing of 100 m with speeds $\dot{x}_i(t) = 30 \text{ km/h}$, $\dot{x}_{i-1}(t) = 10 \text{ km/h}$.
- Scenario 2: On Interstate 90, one vehicle is following another at a spacing of 100 m with speeds $\dot{x}_i(t) = 130 \text{ km/h}$, $\dot{x}_{i-1}(t) = 110 \text{ km/h}$.

The subject driver on Interstate 90 is certainly under a great deal of pressure to maintain safety during car following because, at such a high speed, a moment's lapse of attention would result in a catastrophe. In contrast, the subject driver in downtown Amherst should have peace of mind because, if something goes wrong, he or she can always slam on the brake to stop the vehicle. Hence, our daily driving experiences suggest that the response in scenario 2 be greater than that in scenario 1. However, GM3 predicts no difference because in both scenarios the speed difference is 20 km/h and the spacing is 100 m.

14.1.4 GM4

GM3's inability to differentiate high-speed and low-speed car-following scenarios motivated General Motors researchers to further explore unexplained factors that can be extracted from the sensitivity coefficient. Interestingly, a dimension analysis reveals that the sensitivity coefficient has the same unit

as frequency (i.e., 1/s) in GM1 and GM2 and the same unit as speed (i.e., m/s) in GM3. Since there is a need to explicitly consider speed as a stimulus, it seems ideal to extract speed from the sensitivity coefficient, leaving the remainder as a new, dimensionless coefficient. This gives rise to the fourth-generation model:

$$\ddot{x}_i(t + \tau_i) = \alpha \frac{\dot{x}_i(t + \tau_i)[\dot{x}_{i-1}(t) - \dot{x}_i(t)]}{[\dot{x}_{i-1}(t) - x_i(t)]}. \quad (14.4)$$

14.1.5 GM5

To generalize the results of the above GM models and, as becomes clear later, to facilitate finding “the bridge” between microscopic and macroscopic models, a generic form of GM models is proposed as the fifth model:

$$\ddot{x}_i(t + \tau_i) = \alpha \frac{[\dot{x}_i(t + \tau_i)]^m}{[\dot{x}_{i-1}(t) - x_i(t)]^l} [\dot{x}_{i-1}(t) - \dot{x}_i(t)], \quad (14.5)$$

where x_i , \dot{x}_i , and \ddot{x}_i are the displacement, speed, and acceleration of the subject vehicle i , and similar notation applies to its leader $i - 1$. τ is the perception-reaction time that applies to all drivers, α is a dimensionless sensitivity coefficient, and m and l are speed and spacing exponents, respectively.

14.2 MICROSCOPIC BENCHMARKING

The following segment of code implements GM4, a full-bloom model in the family. At time step j , the displacement, speed, and acceleration of each vehicle are updated:

```

FOR i = 1:I
    v(j,i) = max([0, v(j-1,i) + a(j,i) * dt]);
    d_v = v(j,i-1) - v(j,i);
    x(j,i) = x(j-1,i) + v(j,i) * dt;
    d_x(j,i) = x(j,i-1) - x(j,i);
    delay = ceil(tau_i/dt);
    a(j + delay, i) = alpha * v(j,i) * d_v(j,i) / d_x(j,i);
END

```

where x , v , and a are displacement, speed, and acceleration, respectively; i is the vehicle ID, $i \in \{1, 2, \dots, I\}$; j is the time step, $j \in \{1, 2, \dots, J\}$; τ_i is the perception-reaction time of driver i ; and Δt is the simulation time step.

Microscopic benchmarking refers to the scenario presented in Section 12.3.1 and the benchmarking result of GM4 is plotted in [Figure 14.2](#), which is further elaborated as follows:

- Start-up: the model is unable to start a vehicle from standstill. Therefore, an external logic has to be imposed to assign an initial speed $\dot{x}_i(0)$ to the subject vehicle i . Note that the initial speed $\dot{x}_i(0)$ has to be set at the desired speed v_i . Otherwise, vehicle i will not be able to reach that speed by itself. See [Figure 14.2](#) when $t > 0$ s.
- Speedup: rather than speeding up vehicle i as drivers normally do in the real world, the model predicts a deceleration by driver i even though its leading vehicle $i - 1$ is thousands of meters ahead. See [Figure 14.2](#) when $0 \text{ s} < t < 100 \text{ s}$.
- Free flow: the model predicts that vehicle i is unable to attain the free-flow condition by itself unless it is set to do so by an external logic. As long as it follows a slower leader, the model constantly decelerates the vehicle until it adopts the leader's speed. See [Figure 14.2](#) when $0 \text{ s} < t < 100 \text{ s}$.
- Cutoff: when the third vehicle suddenly takes over as the new leader 40 m ahead at 24 m/s, the model predicts a sudden acceleration, while in the real world drivers may or may not decelerate the vehicle. See [Figure 14.2](#) around $t = 100$ s.
- Following: the model is able to adopt the leader's speed and follow the leader at a reasonable distance. See [Figure 14.2](#) when $100 \text{ s} < t < 200 \text{ s}$.
- Stop and go: the model predicts that vehicle i will gradually but surely collide with its leader while maintaining a speed, regardless of how low the speed is. When the leader resumes motion, vehicle i will be stuck there because of its infinitesimally low speed. See [Figure 14.2](#) when $200 \text{ s} \geq t \leq 300 \text{ s}$.
- Trailing: vehicle i is stuck and fails to catch up with its speeding leader, unless another external logic brings it out of being stuck. See [Figure 14.2](#) when $300 \text{ s} < t < 400 \text{ s}$. However, once the vehicle resumes motion, it will be tempted to catch up with its speeding leader and will adopt the leader's speed. Such an effect is not shown in [Figure 14.2](#).
- Approaching: the simulation fails to be reasonable beyond $t = 300$ s.
- Stopping: the simulation fails to be reasonable beyond $t = 300$ s.

The above benchmarking is based on the set of parameters in [Table 14.1](#), and the outcome may differ for a different set of parameters.

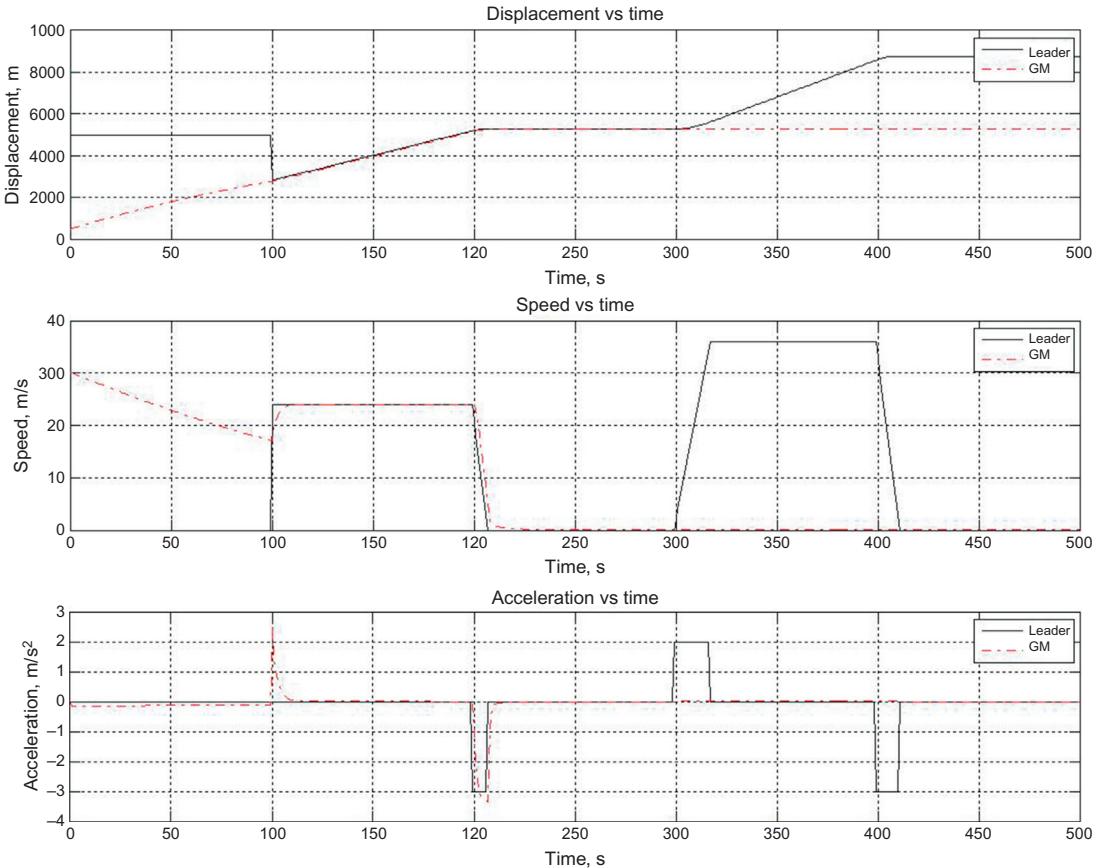


Figure 14.2 Microscopic benchmarking of GM4.

Table 14.1 Microscopic benchmarking parameters of GM4

τ_i	α	—
1.0 s	0.8	—
$x_i(0)$	$\dot{x}_i(0)$	$\ddot{x}_i(0)$
467 m	30 m/s	0 m/s ²

14.3 MICROSCOPIC-MACROSCOPIC BRIDGE

As mentioned in Chapter 13, the relation between microscopic and macroscopic models is always of great interest because such a relation offers a “bridge” to connect microscopic and macroscopic worlds. This section is specifically devoted to such a purpose. It appears that GM5 is ideal to serve as a unifying factor that pulls together some of the microscopic and macroscopic/equilibrium models in the early history of traffic flow theory. For convenience, GM5 is repeated below:

$$\ddot{x}_i(t + \tau_i) = \alpha \frac{[\dot{x}_i(t + \tau_i)]^m}{[x_{i-1}(t) - x_i(t)]^l} [\dot{x}_{i-1}(t) - \dot{x}_i(t)]. \quad (14.6)$$

In addition, those early equilibrium models are listed in [Table 14.2](#).

14.3.1 Greenberg Model

If one chooses $m = 0$ and $l = 1$, GM5 reduces to GM3:

$$\ddot{x}_i(t + \tau_i) = \alpha \frac{[\dot{x}_{i-1}(t) - \dot{x}_i(t)]}{[x_{i-1}(t) - x_i(t)]} \quad (14.7)$$

Table 14.2 Single-regime models

Authors	Model	Parameters
Greenshields [9]	$v = v_f \left(1 - \frac{k}{k_j}\right)$	v_f, k_j
Greenberg [10]	$v = v_m \ln \left(\frac{k_j}{k}\right)$	v_m, k_j
Underwood [11]	$v = v_f e^{-\frac{k}{k_m}}$	v_f, k_m
Drake et al. [12]	$v = v_f e^{-\frac{1}{2} \left(\frac{k}{k_m}\right)^2}$	v_f, k_m
Drew [13]	$v = v_f \left[1 - \left(\frac{k}{k_j}\right)^{n+\frac{1}{2}}\right]$	v_f, k_j, n
Pipes [14] and Munjal [15]	$v = v_f \left[1 - \left(\frac{k}{k_j}\right)^n\right]$	v_f, k_j, n

v_f is free-flow speed, k_j is jam density, v_m is optimal speed, k_m is optimal density, and n is an exponent.

It can be proved that, by integration, this microscopic car-following model can be transformed into the Greenberg model, of which the following is a revised version:

$$\begin{cases} q = v_f k & \text{when } 0 \leq k < k_c, \\ q = v_m \ln \frac{k_j}{k} k & \text{when } k_c \leq k \leq k_j. \end{cases} \quad (14.8)$$

The purpose of the revision to avoid the infinite free-flow speed problem in the Greenberg model based on the following observation. It is known that under zero to light traffic conditions, there is enough room to allow drivers to maintain their desired speed, and hence free-flow speed v_f can be sustained up to a density called the critical density k_c . As traffic density continues to increase, traffic speed begins to drop until it reaches zero, when the density becomes maximum—that is, k_j .

The above flow-density relationship is plotted as the solid curve in [Figure 14.3](#). For easy reference, the underlying speed-density curve is plotted as a dashed line in the background to show how it relates to the flow-density curve.

The above relation between the Greenberg model and GM3 suggests that one might be able to relate some of the existing equilibrium traffic flow models to GM5 by aggregating or integrating this model with varying speed and spacing exponents. The following constitutes some additional examples.

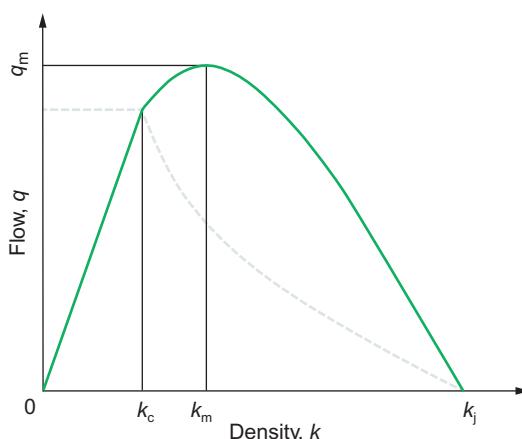


Figure 14.3 Flow-density relationship.

14.3.2 Greenshields Model

Setting $l = 2$ and $m = 0$ in GM5 (Equation 14.5) yields

$$\ddot{x}_i(t + \tau_i) = \alpha \frac{[\dot{x}_{i-1}(t) - \dot{x}_i(t)]}{[x_{i-1}(t) - x_i(t)]^2}, \quad (14.9)$$

where all variables are defined as before. It can be proved that this microscopic model can be transformed into the Greenshields model:

$$v = v_f - \frac{v_f}{k_j} k. \quad (14.10)$$

14.3.3 Underwood Model

Setting $l = 2$ and $m = 1$ in GM5 yields

$$\ddot{x}_i(t + \tau_i) = \alpha \frac{\dot{x}_i(t + \tau_i)[\dot{x}_{i-1}(t) - \dot{x}_i(t)]}{[x_{i-1}(t) - x_i(t)]^2}. \quad (14.11)$$

It can be proved that this microscopic model can be transformed into the Underwood model:

$$v = v_f e^{-k/k_m}. \quad (14.12)$$

14.3.4 Model of Drake et al. (Northwestern Model)

Setting $l = 3$ and $m = 1$ in GM5 yields

$$\ddot{x}_i(t + \tau_i) = \alpha \frac{\dot{x}_i(t + \tau_i)[\dot{x}_{i-1}(t) - \dot{x}_i(t)]}{[x_{i-1}(t) - x_i(t)]^3}. \quad (14.13)$$

It can be proved that this microscopic model can be transformed into the model of Drake et al.:

$$v = v_f e^{-\frac{1}{2} \left(\frac{k}{k_m} \right)^2}. \quad (14.14)$$

14.3.5 Pipes-Munjaj Model

Setting $l = n + 1$ and $m = 0$ in GM5 yields

$$\ddot{x}_i(t + \tau_i) = \alpha \frac{[\dot{x}_{i-1}(t) - \dot{x}_i(t)]}{[x_{i-1}(t) - x_i(t)]^{n+1}}. \quad (14.15)$$

It can be proved that this microscopic model can be transformed into the Pipes-Munjaj model:

$$v = v_f \left[1 - \left(\frac{k}{k_j} \right)^n \right]. \quad (14.16)$$

14.3.6 Drew Model

Since the Drew model and the Pipes-Munjal model are exactly the same except for their exponent, one only needs to replace n with $n + \frac{1}{2}$ in the above derivation to obtain the Drew model. Hence, the Drew model corresponds to GM5 with $l = n + 1.5$ and $m = 0$.

14.3.7 Summary of the Bridge

Summarizing the above, we can draw a diagram that relates the models discussed above to GM5. Figure 14.4 serves such a purpose, with the speed exponent m of GM5 on the horizontal axis and the spacing exponent l of GM5 on the vertical axis. Macroscopic equilibrium models are labeled in red and microscopic car-following models are labeled in blue. Circles on the grid denote models and their corresponding m and l combination in relation to GM5.

The Pipes and Forbes models are actually a special case of GM1:

$$\ddot{x}_i(t + \tau_i) = \alpha[\dot{x}_{i-1}(t) - \dot{x}_i(t)].$$

Integrating both sides yields

$$\dot{x}_i(t + \tau_i) = \alpha[x_{i-1}(t) - x_i(t)] + C = \alpha s_i(t) + C.$$

If one chooses $\alpha = \frac{l_i}{4.47}$ and $C = l_{i-1}$, one obtains the Pipes model, while $\alpha = \tau_i$ and $C = l_i$ leads to the Pipes model.



Figure 14.4 Microscopic-macroscopic bridge.

14.4 MACROSCOPIC BENCHMARKING

Macroscopic benchmarking refers to the scenario presented in Section 12.3.2. Fundamental diagrams implied by GM models and their associated equilibrium models are presented in [Figure 14.5](#) against empirical observations. It can be seen that these fundamental diagrams achieve varying success in fitting empirical data, but none of them fit the data reasonably well in the entire range of density. For example, the Greenshields model overpredicts speed (and hence flow) in the majority of the density range except for the free-flow (i.e., low-density) condition; the Greenberg model has a problem fitting the data under the free-flow condition; the Underwood model, perhaps the best among the models, underestimates speed at low densities and overestimates speed at high densities, and capacity occurs at much lower speed than it ought to; the model of Drake et al. (Northwestern model) has a flow-density curve that is convex in the high-density range; the Drew and Pipes-Munjal models, which are essentially the same but are shown slightly differently to avoid complete overlap, share the same problem as the Greenshields model but to a lesser extent.

The above benchmarking is based on the set of parameters in [Table 14.3](#), and the outcome may differ for a different set of parameters

14.5 LIMITATIONS OF GM MODELS

As seminal work in the early history, GM models spawned and inspired numerous research efforts that have shaped today's traffic flow theory, and thus the importance of this work cannot be underestimated. Meanwhile,

Table 14.3 Macroscopic benchmarking parameters of models associated with GM models

Greenshields	v_f 30 m/s	k_j 1/6 vehicles/m	— —
Greenberg	v_m 10.7 m/s	k_j 1/6 vehicles/m	k_c 0.01 vehicles/m
Underwood	v_f 30 m/s	k_m 0.05 vehicles/m	— —
Drake et al. (Northwestern)	v_f 30 m/s	k_m 0.04 vehicles/m	n 2
Drew	v_f 30 m/s	k_j 1/6 vehicles/m	n 0.1
Pipes and Munjal	v_f 30 m/s	k_j 1/6 vehicles/m	n 0.5

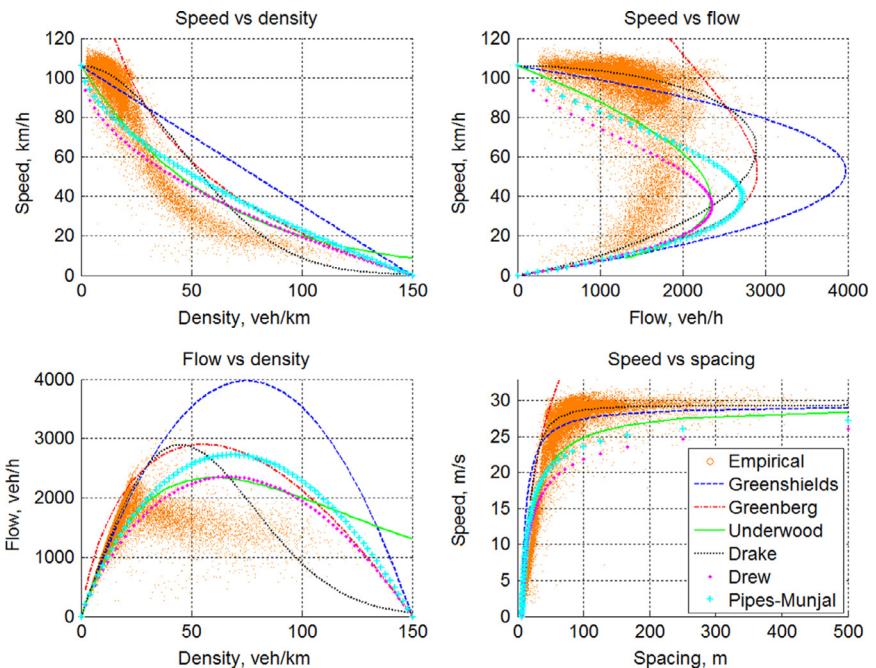


Figure 14.5 Fundamental diagrams implied by GM models and associated models.

GM models suffer from some limitations, which are presented below with use of GM4 as an example.

Universal car following

On the one hand, GM4 is mathematically attractive since it has only one equation that covers all situations. On the other hand, such a one-regime property stipulates universal car following, which is not realistic. For example, the model predicts that a vehicle in Atlanta must be following another vehicle in Boston even though they are over 1000 km apart.

Attraction as a mechanism of motion

If one compares GM4 (Equation 14.4) against Newton's law of universal gravitation (Equation 14.17) and Coulomb's law (Equation 14.18), one finds that they are strikingly similar to each other.

$$F = G \frac{m_1 m_2}{r^2}, \quad (14.17)$$

where F is the force between two masses, G is the gravitational constant, m_1 is the first mass, m_2 is the second mass, and r is the distance between the masses.

$$F = k_e \frac{q_1 q_2}{r^2}, \quad (14.18)$$

where F is the electrostatic force between two point charges (like charges repel each other and opposite charges attract each other), q_1 is the first point charge, q_2 is the second point charge, r is the distance between the two point charges, and k_e is a proportionality constant.

Therefore, GM4 can be interpreted as equivalent to Coulomb's law as follows. Vehicle i will be repelled by its leader $i - 1$ when vehicle i is approaching vehicle $i - 1$ at a higher speed, while vehicle i will be attracted to vehicle $i - 1$ should vehicle i fall behind at a slower speed. Though the first half of the reasoning seems to make some sense, the second half does not. For example, what if the subject vehicle does not have a leader—for example, the first vehicle to enter the highway. Then the subject vehicle could not start because there would be no vehicle to pull it forward. Even if the subject vehicle is following a leader and the gap between them is increasing, it does not feel as if the subject vehicle is attracted to the leader. Actually, the subject vehicle speeds up because one would like to achieve the desired speed.

Slow start

According to GM4, a vehicle at a stopped position is unable to start. This is because the vehicle's speed at the current step ($\dot{x}_i(t) = 0$) determines its acceleration in the next step ($\ddot{x}_i(t+\tau_i) = 0$) (see Equation 14.4). Therefore, the vehicle has to maintain a nonzero speed at any time in order to avoid being trapped. As such, the model fails to apply when a vehicle is stopped by a red light at an intersection or is completely blocked by another vehicle on a highway. Otherwise, the subject vehicle has to slow down to an infinitesimal speed rather than to a complete stop in order to avoid being trapped. When the light turns green or the leading vehicle resumes motion, the subject vehicle will take a long time to get up to speed. This is because the vehicle's infinitesimal speed results in a weak attraction, which is the only mechanism to accelerate the vehicle. [Figure 14.6](#) illustrates such a scenario, where a noticeable gap results between the first vehicle and the second vehicle.

An intimate pair

According to GM4, two vehicles can get arbitrarily close to each other as long as they are traveling at the same speed, which is certainly not true—no

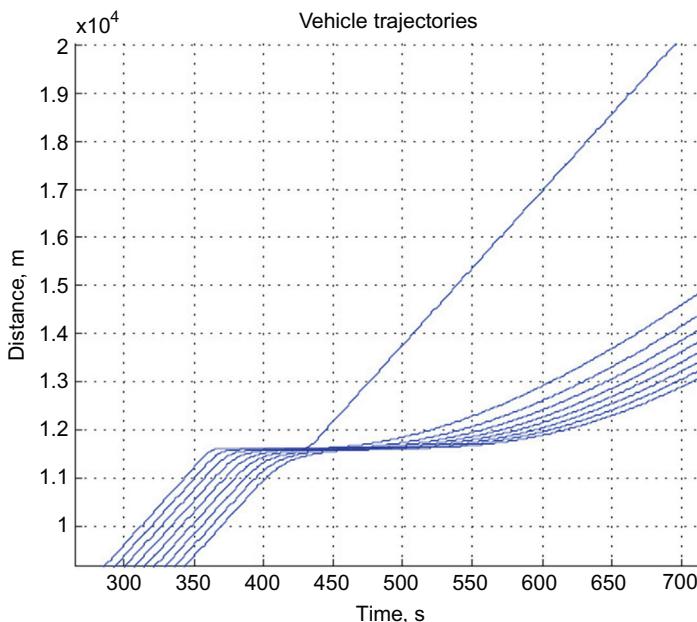


Figure 14.6 GM4 slow start.

one would dare follow another 1 inch apart at 120 km/h! The reason why GM4 allows such a ridiculous car-following distance is because, regardless of how close the two vehicles are, the model predicts no response as long as the two vehicles are moving at the same speed.

PROBLEMS

1. Prove that if one chooses $m = 0$ and $l = 0$, GM5 integrates to the Pipes/Forbes model.
2. Prove that if one chooses $m = 0$ and $l = 1$, GM5 integrates to the Greenberg model.
3. Prove that if one chooses $m = 0$ and $l = 2$, GM5 integrates to the Greenshields model.
4. Prove that if one chooses $m = 1$ and $l = 2$, GM5 integrates to the Underwood model.
5. Prove that if one chooses $m = 1$ and $l = 3$, GM5 integrates to the model of Drake et al. (Northwestern model).
6. Prove that if one chooses $m = 0$ and $l = n + 1$, GM5 integrates to the Pipes-Munjal model.
7. Prove that if one chooses $m = 0$ and $l = n + 1.5$, GM5 integrates to the Drew model.
8. Perform a one-step simulation based on the following conditions: Two cars are traveling in the same lane on a freeway. The length of both vehicles is $l_{i-1} = l_i = 6$ m. Lane change is not considered in this problem. At time t , the leading vehicle $i - 1$ is traveling at a speed of $\dot{x}_{i-1}(t) = 72$ km/h and the following vehicle i is traveling at a speed of $\dot{x}_i(t) = 108$ km/h. The spacing between the two vehicles (measured from front bumper to front bumper) is $s_i(t) = 40$ m. The perception-reaction time of the following driver is $\tau_i = 1.5$ s.
 - a. Use GM1 to predict the deceleration that the following driver will adopt after a perception-reaction time. Assume the sensitivity factor is 0.5 1/s .
 - b. Use GM2 to predict the acceleration that the following driver will adopt after a perception-reaction time.
 - c. Use common sense to decide which sensitivity factor to use.
 - d. Use GM3 to predict the acceleration that the following driver will adopt after a perception-reaction time. Assume the sensitivity factor is 10 m/s .

- e. Use GM4 to predict the acceleration that the following driver will adopt after a perception-reaction time. Assume vehicles keep their speeds unchanged until after a perception-reaction time and the sensitivity factor is 0.5 .
- f. Use GM5 to predict the acceleration that the following driver will adopt after a perception-reaction time. Assume vehicles keep their speeds unchanged until after a perception-reaction time, a sensitivity factor α of 0.5, $l = 2$, and $m = 2$.

CHAPTER 15

Gipps Model

The Pipes, Forbes, and General Motors models introduced in previous chapters are all single-regime models—that is, they have only one equation that applies to the entire driving process and do not consider different driving scenarios or regimes. On the positive side, such models are simple and mathematically attractive. On the flip side, however, their descriptive power is frequently of concern. For example, a driver may encounter different regimes such as start-up, speedup, free flow, cutoff, following, stop and go, trailing, approaching, and stopping. A one-equation model may or may not apply to all regimes. As such, multiregime models might be helpful in capturing different driving scenarios. This chapter introduces a model along this line—the Gipps model.

15.1 MODEL FORMULATION

Just like the Pipes and Forbes models, the Gipps model [57] is derived from a safe driving rule, perhaps a more realistic but conservative one. A driver typically employs a safety rule to evaluate if the current car-following situation is safe. For example,

- The Pipes rule stipulates that at each moment the driver needs to estimate his or her own speed (in miles per hour), divide it by 10, and multiply the quotient by a car length, and the result is the minimum gap that should be maintained. If the actual gap is less, one should fall back; otherwise, it is safe.
- The Forbes rule ensures safe headway. For example, a driver with a perception-reaction time of 3 s can use the following to ensure safety. When the vehicle in front passes a roadside utility pole, start counting “one thousand one, one thousand two, one thousand three.” If the driver passes the pole before the counting is finished, a 3 s headway is not maintained; otherwise, it is safe.

Though the above two safety rules may sound reasonable to a certain degree, rarely do drivers in the real world drive in such a manner. Perhaps a more realistic safety rule is the following. “The driver of the following vehicle selects his speed to ensure that he can bring his vehicle to a safe stop

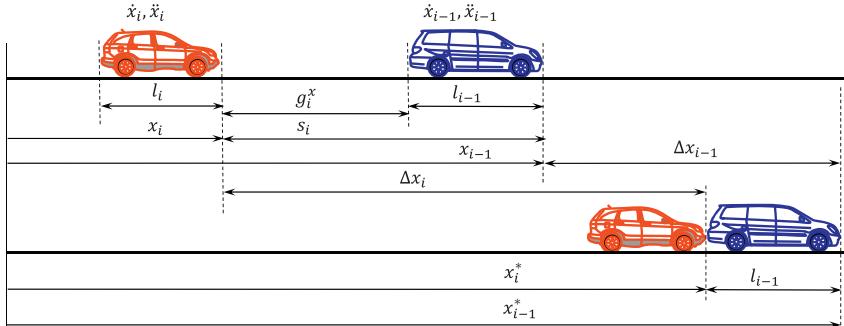


Figure 15.1 Gipps car-following scenario.

should the vehicle ahead come to a sudden stop” [57]. Put another way, at any moment the following driver should leave enough safe distance in front such that in case the leading vehicle commences an emergency brake, the subject driver has time to respond and decelerate to a stop behind the leading vehicle without a collision. The Gipps car-following model is based on such an assumption, and the scenario is depicted in Figure 15.1.

At time t , vehicle i is located at $x_i(t)$ and the leading vehicle $i - 1$ is at $x_{i-1}(t)$. At this moment, vehicle $i - 1$ at speed $\dot{x}_{i-1}(t)$ commences an emergency brake at a rate of B_{i-1} . Alerted by the braking light in front, driver i at speed $\dot{x}_i(t)$ goes through a perception-reaction process of duration τ_i , trying to understand the situation, evaluate potential options, and then decides to brake as well at a *tolerable* rate of b_i . Hence, the vehicle starts to decelerate from $\dot{x}_i(t + \tau_i)$ to a stop, with the most adverse situation being stopped right after vehicle $i - 1$.

Therefore, the distance traveled by vehicle $i - 1$ during its emergency brake is $-\frac{\dot{x}_{i-1}^2(t)}{2B_{i-1}}$ since B_{i-1} is negative, so the vehicle stops at location

$$x_{i-1}^* = x_{i-1}(t) - \frac{\dot{x}_{i-1}^2(t)}{2B_{i-1}}. \quad (15.1)$$

Meanwhile, vehicle i travels a distance of $\frac{\dot{x}_i(t) + \dot{x}_i(t + \tau_i)}{2}\tau_i$ during the perception-reaction time and then travels a braking distance of $-\frac{\dot{x}_i(t + \tau_i)^2}{2b_i}$. Hence, the vehicle stops at location

$$x_i^* = x_i(t) + \frac{\dot{x}_i(t) + \dot{x}_i(t + \tau_i)}{2}\tau_i - \frac{\dot{x}_i^2(t + \tau_i)}{2b_i}. \quad (15.2)$$

To be conservative, Gipps added one more buffer space term in the above equation:

$$x_i^* = x_i(t) + \frac{\dot{x}_i(t) + \dot{x}_i(t + \tau_i)}{2} \tau_i + \dot{x}_i(t + \tau_i) \theta - \frac{\dot{x}_i^2(t + \tau_i)}{2b_i}, \quad (15.3)$$

where θ is an extra buffer time appended to the perception-reaction time. To ensure safety, the following relationship must hold:

$$x_{i-1}^* - l_{i-1} \geq x_i^*. \quad (15.4)$$

Plugging in everything, we find the above inequality translates to

$$x_{i-1}(t) - \frac{\dot{x}_{i-1}^2(t)}{2B_{i-1}} - l_{i-1} \geq x_i(t) + \frac{\dot{x}_i(t) + \dot{x}_i(t + \tau_i)}{2} \tau_i + \dot{x}_i(t + \tau_i) \theta - \frac{\dot{x}_i^2(t + \tau_i)}{2b_i}. \quad (15.5)$$

Note that the actual spacing is $s_i(t) = x_{i-1}(t) - x_i(t)$, so the safe spacing is

$$s_i(t) \geq \frac{\dot{x}_i(t) + \dot{x}_i(t + \tau_i)}{2} \tau_i + \dot{x}_i(t + \tau_i) \theta - \frac{\dot{x}_i^2(t + \tau_i)}{2b_i} + \frac{\dot{x}_{i-1}^2(t)}{2B_{i-1}} + l_{i-1}. \quad (15.6)$$

Though the above inequality can serve the need of a safety check, the driver needs a basis to determine what to do next in order to achieve the safety goal. Hence, it is necessary to identify which variables in the above inequality are the inputs and which variable is the output. Our daily driving experience suggests the following: $s_i(t)$ and l_{i-1} can be visually estimated; $\dot{x}_i(t)$ and $\dot{x}_{i-1}(t)$ are measurable from the speedometer or motion relative to the roadside; τ_i , θ , b_i , and B_{i-1} are internal to the driver and hence are implicitly known. As such, these variables can be treated as the inputs, while the only output in the above inequality is $\dot{x}_i(t + \tau_i)$, which is the target speed that the driver tries to achieve next. Therefore, finding the output translates to solving the following quadratic inequality:

$$-\frac{1}{2b_i} \dot{x}_i^2(t + \tau_i) + \left(\frac{\tau_i}{2} + \theta \right) \dot{x}_i(t + \tau_i) + \frac{\dot{x}_i(t) \tau_i}{2} + \frac{\dot{x}_{i-1}^2(t)}{2B_{i-1}} + l_{i-1} - s_i(t) \leq 0. \quad (15.7)$$

The roots of the above quadratic equation are

$$\begin{aligned} \dot{x}_i(t + \tau_i) &= b_i \left(\frac{\tau_i}{2} + \theta \right) \\ &\pm \sqrt{b_i^2 \left(\frac{\tau_i}{2} + \theta \right)^2 - b_i \left[-\dot{x}_i(t) \tau_i - \frac{\dot{x}_{i-1}^2(t)}{B_{i-1}} - 2l_{i-1} + 2s_i(t) \right]}. \end{aligned} \quad (15.8)$$

Let $\theta = \tau_i/2$ as suggested by Gipps. Then,

$$\dot{x}_i(t + \tau_i) = -b_i \tau_i \pm \sqrt{b_i^2 \tau_i^2 - b_i[-\dot{x}_i(t) \tau_i - \frac{\dot{x}_{i-1}^2(t)}{B_{i-1}} - 2l_{i-1} + 2s_i(t)]}. \quad (15.9)$$

Consider the signs of the roots and that speed is a positive value, then the solution to inequality (15.7) is

$$0 \leq \dot{x}_i(t + \tau_i) \leq -b_i \tau_i + \sqrt{b_i^2 \tau_i^2 - b_i[-\dot{x}_i(t) \tau_i - \frac{\dot{x}_{i-1}^2(t)}{B_{i-1}} - 2l_{i-1} + 2s_i(t)]}. \quad (15.10)$$

The above derivation has formulated the Gipps model in the car-following regime. In the free-flow regime—that is,, vehicle i is not blocked by a leading vehicle, Gipps suggests the following speed choice:

$$0 \leq \dot{x}_i(t + \tau_i) \leq \dot{x}_i(t) + 2.5A_i \tau_i \left(1 - \frac{\dot{x}_i(t)}{v_i}\right) \sqrt{0.025 + \frac{\dot{x}_i(t)}{v_i}}, \quad (15.11)$$

where A_i is the maximum acceleration that driver i is willing to apply and v_i is the desirable speed that driver i is willing to travel at whenever possible. Unlike inequality (15.10), which is derived from a safety rule, this speed choice model is an empirical one obtained from fitting vehicle experimental data. It basically accelerates/decelerates the vehicle to the desirable speed without causing oscillation.

The above two speed choices may cause a little confusion when they are applied to vehicle control because one has to constantly decide for one of them. For example, one should choose (15.10) in the case of car following and (15.11) in the case of free flow. However, what is the cutoff point between free flow and car following? To resolve this confusion, Gipps suggests that there is no need to make a distinction. Under any situation, one just needs to compute the two speeds and choose the lower one—that is,

$$\dot{x}_i(t + \tau_i) = \min \begin{cases} \dot{x}_i(t) + 2.5A_i \tau_i \left(1 - \frac{\dot{x}_i(t)}{v_i}\right) \sqrt{0.025 + \frac{\dot{x}_i(t)}{v_i}} & (\text{free flow}), \\ -b_i \tau_i + \sqrt{b_i^2 \tau_i^2 - b_i[\dot{x}_i(t) \tau_i - \frac{\dot{x}_{i-1}^2(t)}{B_{i-1}} + 2l_{i-1} - 2s_i(t)]} & (\text{car following}). \end{cases} \quad (15.12)$$

15.2 PROPERTIES OF THE GIPPS MODEL

As usual, the macroscopic property of the Gipps model under equilibrium conditions is of primary interest. To simplify the analysis, the Gipps safety rule presented in (15.6) can be simplified as follows if one ignores the speed change during the perception-reaction process and the additional buffer time θ :

$$s_i(t) \geq \dot{x}_i(t)\tau_i - \frac{\dot{x}_i^2(t)}{2b_i} + \frac{\dot{x}_{i-1}^2(t)}{2B_{i-1}} + l_{i-1}. \quad (15.13)$$

Setting both sides equal and rearranging terms yields

$$s_i(t) = \dot{x}_i(t)\tau_i - \frac{\dot{x}_i^2(t)}{2b_i} + \frac{\dot{x}_{i-1}^2(t)}{2B_{i-1}} + l_{i-1}. \quad (15.14)$$

Under equilibrium conditions, the above car-following model leads to the following speed-density relationship:

$$\frac{1}{k} = \left(-\frac{1}{2b} + \frac{1}{2B} \right) v^2 + \tau v + l \quad (15.15)$$

or

$$\frac{1}{k} = \gamma v^2 + \tau v + l, \quad (15.16)$$

where k is traffic density, $\gamma = -\frac{1}{2b} + \frac{1}{2B}$, $b < 0$ is the average tolerable braking rate, $B < 0$ is the average emergency braking rate, v is the average traffic speed, τ is the average perception-reaction time, and l is the average nominal vehicle length. The corresponding flow-speed relationship is

$$q = \frac{v}{\gamma v^2 + \tau v + l}. \quad (15.17)$$

To find the capacity, one takes the first derivative of flow q with respect to v and sets the result to zero:

$$\frac{dq}{dv} \Big|_{v_m} = -\frac{\gamma - \frac{l}{v^2}}{(\gamma v + \tau' + \frac{l}{v})^2} \Big|_{v_m} = 0. \quad (15.18)$$

Solving the equation yields

$$v_m = \sqrt{\frac{l}{\gamma}}, \quad (15.19)$$

and correspondingly,

$$q_m = \frac{1}{2\sqrt{\gamma l} + \tau}. \quad (15.20)$$

From a check of the second derivative of q at v_m , it turns out that q_m is indeed the maximum value of q .

15.3 BENCHMARKING

The microscopic benchmarking refers to the scenario presented in Section 12.3.1 and the macroscopic benchmarking refers to the scenario presented in Section 12.3.2.

15.3.1 Microscopic Benchmarking

The benchmarking result of the Gipps model is plotted in [Figure 15.2](#). The performance of the Gipps model is summarized as follows:

- Start-up: the model is able to start the vehicle up from standstill. See [Figure 15.2](#) when $t > 0$ s.
- Speedup: the model is able to speed the vehicle up realistically to its desired speed. See [Figure 15.2](#) when $0 \text{ s} < t < 100$ s.
- Free flow: the model is able to reach and settle at the desired speed under free-flow conditions. See [Figure 15.2](#) when $0 \text{ s} < t < 100$ s.
- Cutoff: the model overdecelerates slightly, which causes a small oscillation in speed, but in general the model retains control and responds reasonably when a vehicle cuts in in front. See [Figure 15.2](#) around $t = 100$ s.
- Following: the model is able to adopt the leader's speed and follow the leader at a reasonable distance. See [Figure 15.2](#) when $100 \text{ s} < t < 200$ s.
- Stop and go: the model is able to stop the vehicle safely behind its leader and start the vehicle moving when the leader departs. See [Figure 15.2](#) when $200 \text{ s} \geq t \leq 300$ s.
- Trailing: the model is able to speed up normally without being tempted to speed up by its speeding leader. See [Figure 15.2](#) when $300 \text{ s} < t < 400$ s.
- Approaching: the model is able to decelerate properly when approaching a stationary vehicle at a distance. See [Figure 15.2](#) when $400 \text{ s} \geq t < 420$ s.
- Stopping: the model is able to stop the vehicle safely behind the stationary vehicle. See [Figure 15.2](#) when $t \geq 420$ s.

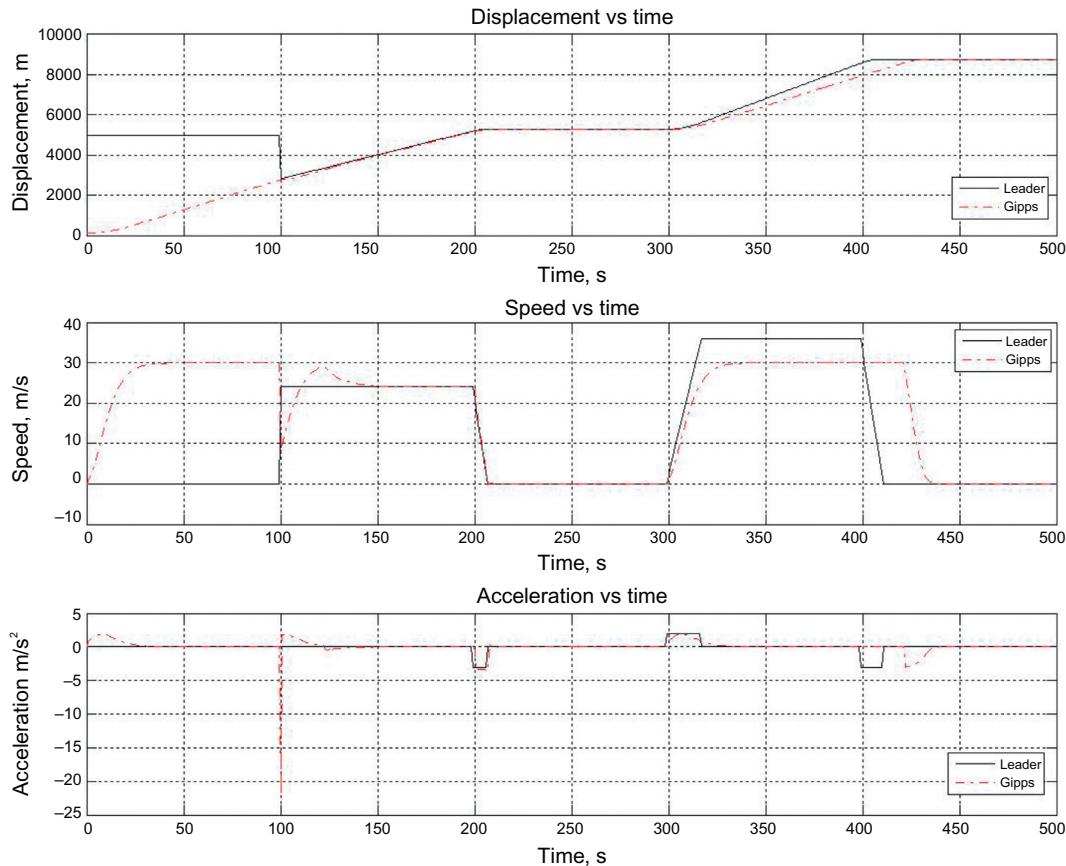


Figure 15.2 Microscopic benchmarking of the Gipps model.

Table 15.1 Microscopic benchmarking parameters of the Gipps model

I_i	v_i	τ_i	b_i	
6 m	30 m/s	1.0 s	-3.4 m/s^2	
A_i	B_{i-1}	$x_i(0)$	$\dot{x}_i(0)$	$\ddot{x}_i(0)$
1.7 m/s^2	-6.0 m/s^2	120 m	0 m/s	0 m/s^2

The above benchmarking is based on the set of parameters in [Table 15.1](#), and the outcome may differ for a different set of parameters.

15.3.2 Macroscopic Benchmarking

The fundamental diagram implied by the Gipps model is plotted in [Figure 15.3](#) against empirical observations. The model parameters are set the same as suggested in the original paper. It can be seen that the model fits empirical

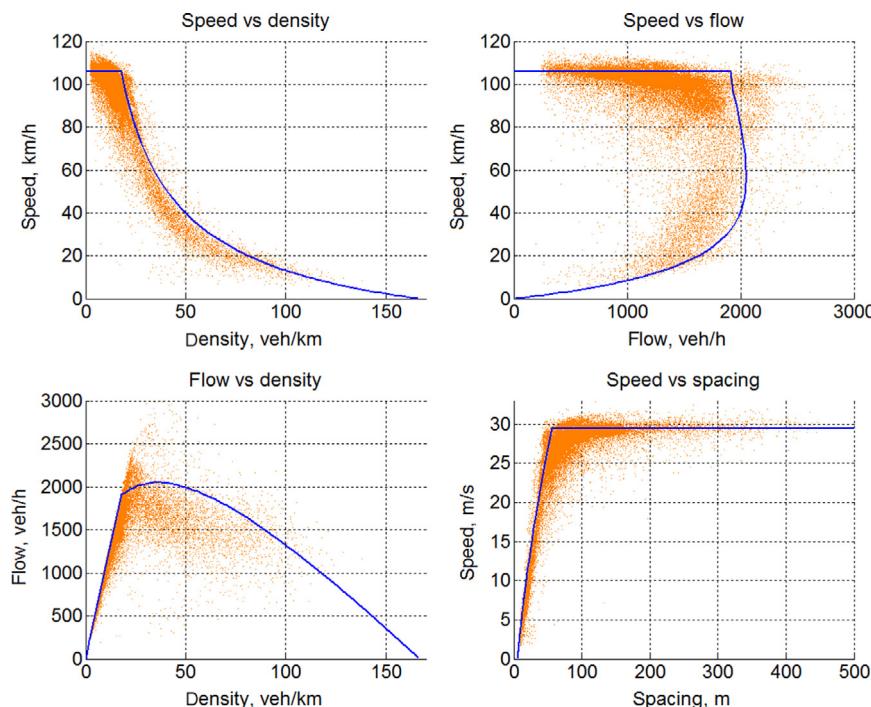


Figure 15.3 Fundamental diagram implied by the Gipps model.

Table 15.2 Macroscopic benchmarking parameters of the Gipps model

<i>b</i>	<i>B</i>	<i>τ</i>	<i>l</i>
-3.0 m/s^2	-3.5 m/s^2	1 s	6.5 m

data reasonably well except for free-flow conditions (i.e., in the low-density range). In addition, the speed at capacity predicted by the Gipps model is much lower than it should be.

The above benchmarking is based on the set of parameters in [Table 15.2](#), and the outcome may differ for a different set of parameters.

PROBLEMS

1. Compare the Forbes model and the Gipps model and explain the difference in their modeling philosophy.
2. The following figure is the result of macroscopic benchmarking of the Gipps model with the following model formulation and parameters:

$$\frac{1}{k} = \left(-\frac{1}{2b} + \frac{1}{2B} \right) v^2 + \tau v + l,$$

where tolerable deceleration $b = -3.0 \text{ m/s}^2$, emergency braking rate $B = -3.5 \text{ m/s}^2$, perception-reaction time $\tau = 1.0 \text{ s}$ and nominal vehicle length $l = 6.5 \text{ m}$.

Find the capacity condition (q_m, k_m, v_m) of the model and comment on your result.

3. Perform a one-step simulation based on the following conditions: Two cars are traveling in the same lane on a freeway. The length of both vehicles is $l_{i-1} = l_i = 6 \text{ m}$. Lane change is not considered in this problem. At time t , the leading vehicle $i - 1$ is traveling at a speed of $\dot{x}_{i-1}(t) = 72 \text{ km/h}$ and the following vehicle i is traveling at a speed of $\dot{x}_i(t) = 108 \text{ km/h}$. The spacing between the two vehicles (measured from front bumper to front bumper) is $s_i(t) = 40 \text{ m}$. The perception-reaction time of the following driver is $\tau_i = 1.5 \text{ s}$. Use the Gipps model to predict the speed that the following driver will adopt after a perception-reaction time. Assume the model parameters take the same values as those in [Section 15.3.1](#).

4. Building on the above problem, at time t , a third vehicle at speed 108 km/h cuts in between the two vehicles. The spacing between the third vehicle and the following vehicle is 15 m. Use the Gipps model to compute the speed that the following driver needs to adopt after a perception-reaction time.

CHAPTER 16

More Single-Regime Models

This chapter presents a few more car-following models, including the Newell nonlinear model, the Newell simplified model, the intelligent driver model (IDM), and the Van Aerde model. Except for the Newell simplified model, these models not only capture the essence of car-following behavior but also aggregate to sound macroscopic behavior.

16.1 NEWELL NONLINEAR MODEL

Newell actually proposed two car-following models, one in 1961, which will be referred to as the “Newell nonlinear” car-following model [58], and the other in 2002, which will be referred to as the “Newell simplified” car-following model [59] hereafter. The Newell nonlinear car-following model takes the following form:

$$\dot{x}_i(t + \tau_i) = v_i(1 - e^{-\frac{\lambda_i}{v_i}(s_i(t) - l_i)}), \quad (16.1)$$

where $\dot{x}_i(t)$ is the speed of the vehicle with ID i at time t , τ_i is driver i 's perception-reaction time, v_i is driver i 's desired speed, λ_i is a parameter associated with driver i (i.e., the slope of driver i 's speed-spacing curve evaluated at $\dot{x}_i = 0$), $s_i = x_{i-1} - x_i$ is the spacing between vehicle i and its leader $i - 1$, and l_i is the minimum value of s_i , which can be viewed as the nominal vehicle length. Note that, in microscopic modeling, the driver and his/her vehicle is considered as a single unit and treated as a particle. Therefore, the “driver”, the “vehicle”, the “unit”, and the “particle” are the same thing and used interchangeably according to the context throughout the book.

16.1.1 Properties of the Newell Nonlinear Model

Newell acknowledged that “no motivation for this choice is proposed other than the claim that it has approximately the correct shape and is reasonably simple.” This acknowledgment seems to tell us two things:

1. Unlike the Pipes, Forbes, and Gipps models, which are derived from driving experiences such as safety rules, this model does not seem to be

based on driving experiences, but seems to be based rather on a discovery after some contemplation and empirical studies.

2. If there were something behind the contemplation, it might have been *the correct shape*—the model leads to an equilibrium speed-density curve that resembles field observations.

Under equilibrium conditions, Equation 16.1 reduces to the following speed-density relationship:

$$\nu = \nu_f \left(1 - e^{-\frac{\lambda}{\nu_f} \left(\frac{1}{k} - \frac{1}{k_j} \right)} \right). \quad (16.2)$$

where ν is traffic speed, which is aggregated from vehicle speed \dot{x}_i , ν_f is free-flow speed, which is aggregated from ν_i , λ is a parameter aggregated from λ_i , k is traffic density, which is the reciprocal of average spacing s , which, in turn, is aggregated from spacing s_i , and k_j is jam density, which is the reciprocal of average vehicle length l , which, in turn, is aggregated from nominal vehicle length l_i .

16.1.2 Benchmarking

Microscopic benchmarking refers to the scenario presented in Section 12.3.1 and macroscopic benchmarking refers to the scenario presented in Section 12.3.2.

Microscopic Benchmarking

The benchmarking result of the Newell nonlinear model is plotted in Figure 16.1. The performance of the Newell nonlinear model is summarized as follows:

- Start-up: The model is able to start a vehicle up from standstill. See Figure 16.1 when $t > 0$ s.
- Speedup: The model allows the vehicle speed to jump from 0 to 30 m/s in one time step, resulting in an acceleration of 30 m/s^2 . This is unrealistic, so an external logic has to be imposed to limit the maximum acceleration. Note that simply setting a limiting acceleration would result in an unrealistic acceleration profile (e.g., the vehicle may attain maximum acceleration at high speeds). Therefore, a more realistic acceleration logic is necessary. However, with this addition, the Newell nonlinear model ceases to be a steady-state model, and instead becomes a dynamic model. See Figure 16.1 when $0 \text{ s} < t < 100$ s.
- Free flow: The model is able to reach and settle at the desired speed under free-flow conditions. See Figure 16.1 when $0 \text{ s} < t < 100$ s.

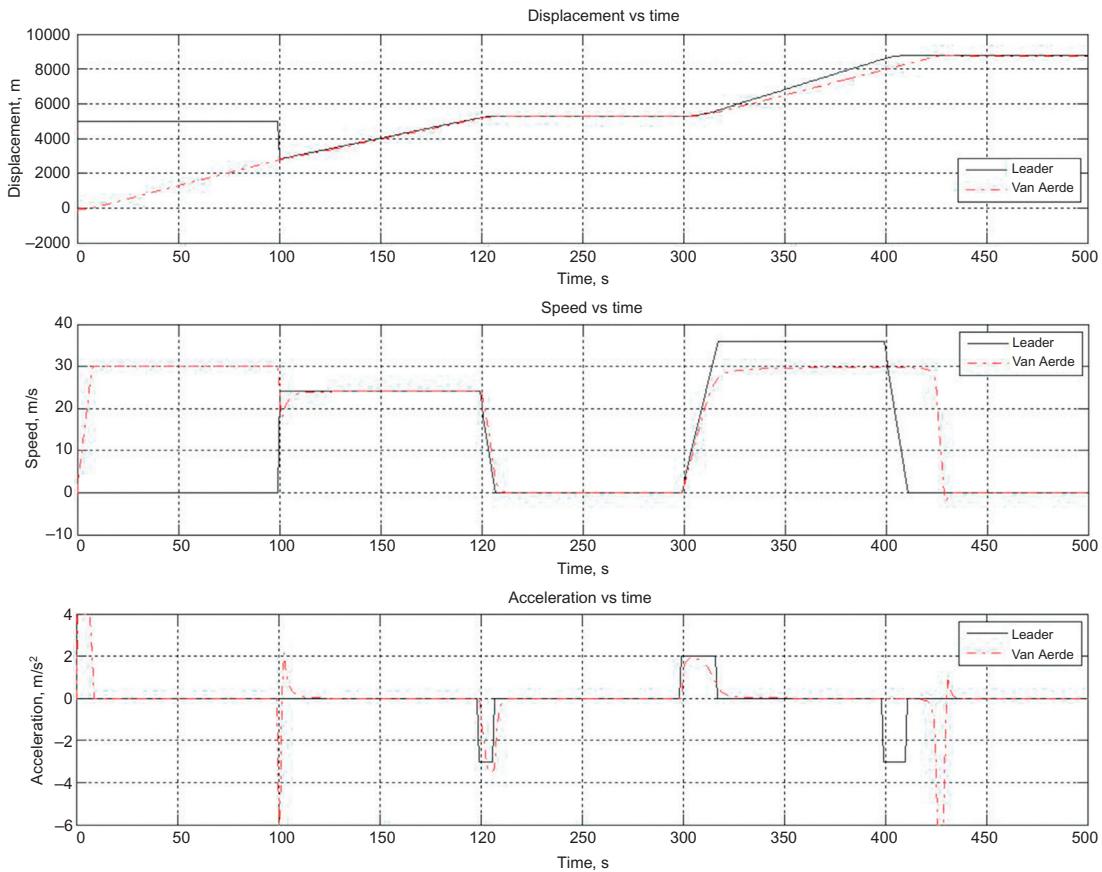


Figure 16.1 Microscopic benchmarking of the Newell nonlinear model.

Table 16.1 Microscopic benchmarking parameters of the Newell nonlinear model

l_i	v_i	τ_i	λ	—
6 m	30 m/s	1.0 s	7.9	—
A_i	B_i	$x_i(0)$	$\dot{x}_i(0)$	$\ddot{x}_i(0)$
4.0 m/s ²	6.0 m/s ²	-97 m	0 m/s	0 m/s ²

- Cutoff: By itself, the Newell nonlinear model would predict a deceleration of about -184.6 m/s^2 when the third vehicle cuts in and an acceleration of 182.9 m/s^2 in the next time step. This is a very unrealistic jerking, so an external logic has to be imposed to limit the maximum acceleration and deceleration. Hence, the same argument as for speedup applies here. See [Figure 16.1](#) around $t = 100 \text{ s}$ after these external conditions have been incorporated.
- Following: The model is able to adopt the leader's speed and follow the leader at a reasonable distance. See [Figure 16.1](#) when $100 \text{ s} < t < 200 \text{ s}$.
- Stop and go: The model is able to stop the vehicle safely behind its leader and start the vehicle moving when the leader departs. See [Figure 16.1](#) when $200 \text{ s} \geq t \leq 300 \text{ s}$.
- Trailing: The model is able to speed up normally without being tempted to speed up by its speeding leader. See [Figure 16.1](#) when $300 \text{ s} < t < 400 \text{ s}$.
- Approaching: With the above external logic on limiting deceleration, the model is able to decelerate properly when approaching a stationary vehicle. See [Figure 16.1](#) when $400 \text{ s} \geq t < 420 \text{ s}$.
- Stopping: The model is able to stop the vehicle safely behind the stationary vehicle. See [Figure 16.1](#) when $t \geq 420 \text{ s}$.

The above benchmarking is based on the set of parameters in [Table 16.1](#), and the outcome may differ for a different set of parameters.

Macroscopic Benchmarking

The fundamental diagram implied by the Newell nonlinear model is presented in [Figure 16.2](#), where the model parameters are adopted from Newell's original paper.

The Newell nonlinear model indeed exhibits the correct shape that resembles field observations in the entire density range, as claimed by Newell. First, the model meets the boundary conditions at $(k = 0, v = v_f)$ and $(k = k_j, v = 0)$. Second, the flow-density exhibits a concave shape, and the fitting quality is reasonably good given that only three parameters are employed.

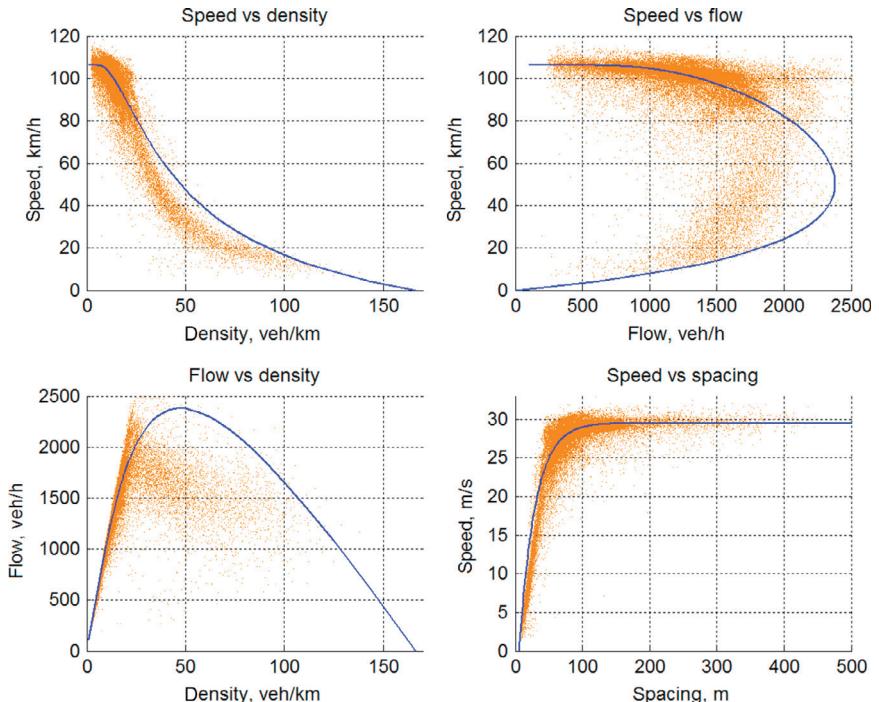


Figure 16.2 Fundamental diagram implied by the Newell nonlinear model.

Table 16.2 Macroscopic benchmarking parameters of the Newell nonlinear model

v_f	k_j	λ
29.5 m/s	0.2 vehicles/m	0.8

The above benchmarking is based on the set of parameters in [Table 16.2](#), and the outcome may differ for different set of parameters.

16.2 NEWELL SIMPLIFIED MODEL

After about 40 years, Newell published a simplified car-following model [59]. This is indeed a very simple model because one does not need to worry about safety rules, speed choices, and acceleration responses. What one needs to do is simply to translate the leading vehicle's trajectory. For example, if vehicle $i-1$'s trajectory $x_{i-1}(t)$ is given in the right panel in [Figure 16.3](#), vehicle i 's trajectory can be directly determined by the following equation:

$$x_i(t + \tau_i) = x_{i-1}(t) - l_i. \quad (16.3)$$

Graphically, this means translating trajectory $x_{i-1}(t)$ to the right by a horizontal distance of τ_i and then downward by a vertical distance of l_i —that is, one can squeeze a rectangle with dimensions $\tau_i \times l_i$ between the two trajectories. From the speed-spacing relationship in the right panel in [Figure 16.3](#), it becomes clear that the physical meaning of l_i is the minimum value of the spacing—that is, the nominal vehicle length—and τ_i is the reciprocal of the tangent to the speed-spacing relationship drawn at point $(0, l_i)$. Evidence shows that τ_i can most likely be interpreted as the perception-reaction time of driver i . [Figure 16.3](#) also reveals that the spacing between the two vehicles is

$$s_i(t) = x_{i-1}(t) - x_i(t). \quad (16.4)$$

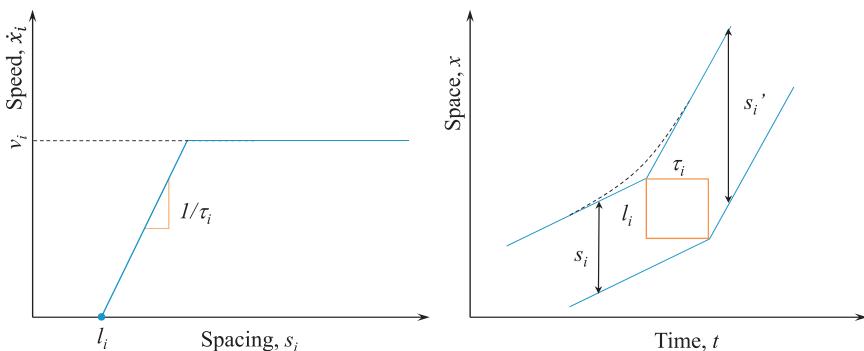


Figure 16.3 Newell simplified car-following model.

In addition, the locations of vehicle i at time t and $t + \tau_i$ can be related as

$$x_i(t + \tau_i) = x_i(t) + \dot{x}_i(t)\tau_i. \quad (16.5)$$

Combining the above three equations, we get

$$s_i(t) = \dot{x}_i(t)\tau_i + l_i. \quad (16.6)$$

This is the same as the Pipes/Forbes model (by taking the minimum spacing), which, in turn, is equivalent to GM1. Since the Newell simplified model is essentially the Pipes/Forbes model, the properties and benchmarking of the latter apply to the former.

16.3 INTELLIGENT DRIVER MODEL

The intelligent driver model (IDM) [60, 61] is expressed as a superposition of the follower i 's acceleration term and a deceleration term which depends on the desired spacing s_i^* :

$$\ddot{x}_i(t + \tau_i) = A_i \left[1 - \left(\frac{\dot{x}_i}{v_i} \right)^\delta - \left(\frac{s_i^*}{s_i} \right)^2 \right], \quad (16.7)$$

where \ddot{x}_i is driver i 's acceleration, A_i is driver i 's maximum acceleration when starting from standstill, δ is the acceleration exponent, $s_i = x_{i-1} - x_i$ is the spacing between vehicle i and its leader $i - 1$, and the desired spacing s_i^* is a function of speed \dot{x}_i and relative speed $(\dot{x}_i - \dot{x}_{i-1})$:

$$s_i^* = s_0 + s_1 \sqrt{\frac{\dot{x}_i}{v_i}} + T_i \dot{x}_i + \frac{\dot{x}_i [\dot{x}_i - \dot{x}_{i-1}]}{2\sqrt{g_i b_i}}, \quad (16.8)$$

where s_0 , s_1 , and T_i are parameters.

16.3.1 Properties of the IDM

Under equilibrium conditions, Equation 16.7 reduces to the following density-speed relationship:

$$k = \frac{1}{(s_0 + vT) \left[1 - \left(\frac{v}{v_f} \right)^\delta \right]^{-1/2}}. \quad (16.9)$$

If one further assumes that $s_0 = s_1 = 0$ and $\delta = 1$, a special case of Equation 16.9 results:

$$\nu = \frac{(s - L)^2}{2\nu_f T^2} \left[-1 + \sqrt{1 + \frac{4T^2\nu_f^2}{(s - L)^2}} \right], \quad (16.10)$$

where T is the average safe time headway, $s = 1/k$ is the average spacing, and k is traffic density.

16.3.2 Benchmarking

Microscopic benchmarking refers to the scenario presented in Section 12.3.1 and macroscopic benchmarking refers to the scenario presented in Section 12.3.2.

Microscopic Benchmarking

The benchmarking result of the IDM is plotted in [Figure 16.4](#). The performance of the IDM is summarized as follows:

- Start-up: The model is able to start the vehicle up from standstill. See [Figure 16.4](#) when $t > 0$ s.
- Speedup: The model is able to speed the vehicle up realistically to its desired speed. See [Figure 16.4](#) when $0 \text{ s} < t < 100$ s.
- Free flow: The model is able to reach and settle at the desired speed under free-flow conditions. See [Figure 16.4](#) when $0 \text{ s} < t < 100$ s.
- Cutoff: The model retains control and responds reasonably when a vehicle cuts in in front. See [Figure 16.4](#) around $t = 100$ s.
- Following: The model is able to adopt the leader's speed and follow the leader at a reasonable distance. See [Figure 16.4](#) when $100 \text{ s} < t < 200$ s.
- Stop and go: The model exhibits some oscillation in acceleration, stopping behind the leading vehicle. The model is able to start moving when the leader departs. See [Figure 16.4](#) when $200 \text{ s} \geq t \leq 300$ s.
- Trailing: The model is able to speed up normally without being tempted to speed up by its speeding leader. See [Figure 16.4](#) when $300 \text{ s} < t < 400$ s.
- Approaching: The model is able to decelerate properly when approaching a stationary vehicle. See [Figure 16.4](#) when $400 \text{ s} \geq t < 420$ s.
- Stopping: The model is able to stop the vehicle safely behind the stationary vehicle. See [Figure 16.4](#) when $t \geq 420$ s.

The above benchmarking is based on the set of parameters in [Table 16.3](#), and the outcome may differ for a different set of parameters.

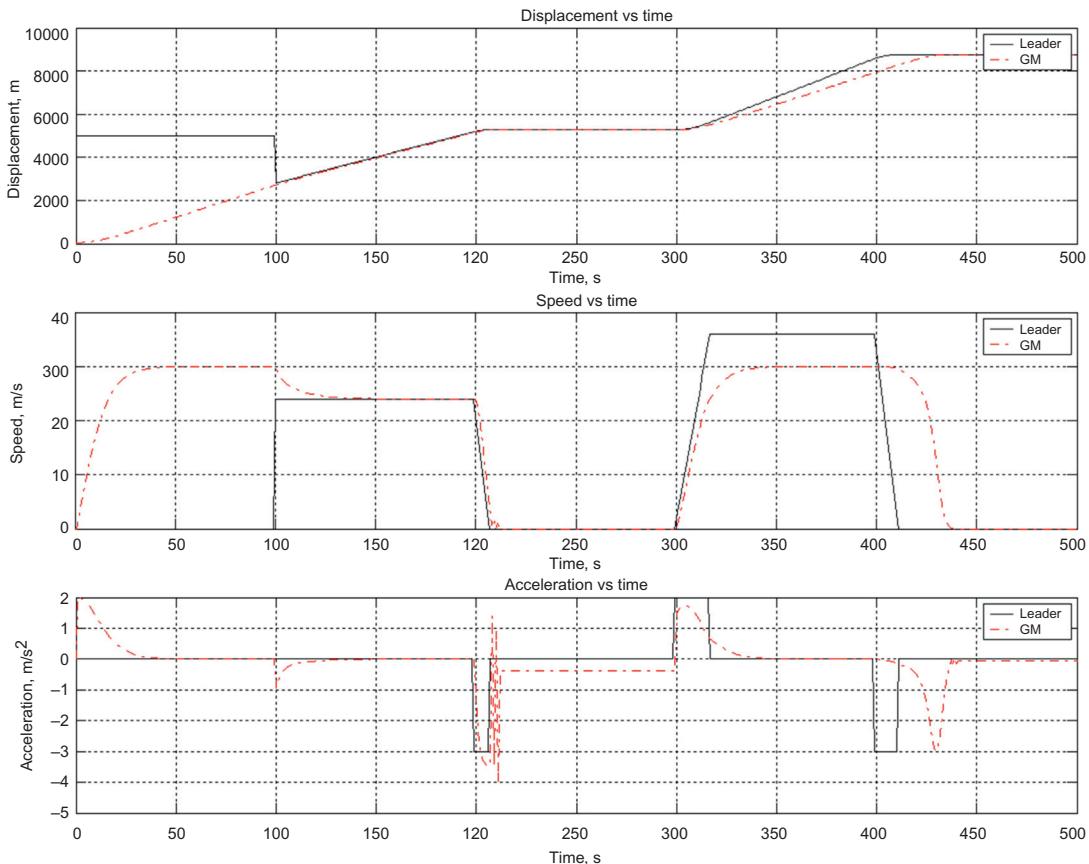


Figure 16.4 Microscopic benchmarking of the IDM.

Table 16.3 Microscopic benchmarking parameters of the IDM

l_i	v_i	τ_i	δ	s_0
6 m	30 m/s	1.0 s	2	2 m
A_i	b_i	$x_i(0)$	$\dot{x}_i(0)$	$\ddot{x}_i(0)$
2.0 m/s ²	4.0 m/s ²	39.5 m	0 m/s	0 m/s ²

Table 16.4 Macroscopic benchmarking parameters of the IDM

v_f	T	δ	s_0
29.5 m/s	1.7 s	15	4 m

Macroscopic Benchmarking

The fundamental diagram implied by the IDM, in particular Equation 16.9, is presented in Figure 16.5. The model employs four parameters and exhibits a desirable shape with good fitting quality.

The above benchmarking is based on the set of parameters in Table 16.4, and the outcome may differ for a different set of parameters.

16.4 VAN AERDE MODEL

The Van Aerde car-following model [62, 63] combines the Pipes model [52] and the Greenshields model [9] into a single equation:

$$s_i = c_1 + c_3 \dot{x}_i + c_2 / (v_f - \dot{x}_i), \quad (16.11)$$

where

$$\begin{cases} c_1 &= \frac{v_f}{k_j v_m^2} (2v_m - v_f), \\ c_2 &= \frac{v_f}{k_j v_m^2} (v_f - v_m)^2, \\ c_3 &= \frac{1}{q_m} - \frac{v_f}{k_j v_m^2}, \end{cases} \quad (16.12)$$

where v_f is the free-flow speed of the roadway facility, k_j is the jam density, and v_m is the optimal speed at capacity q_m .

16.4.1 Properties of the Van Aerde Model

Under equilibrium conditions, Equation 16.11 reduces to the following density-speed relationship:

$$k = \frac{1}{c_1 + c_3 v + c_2 / (v_f - v)}, \quad (16.13)$$

where all variables are as defined before.

16.4.2 Benchmarking

Microscopic Benchmarking

The benchmarking result of the Van Aerde model is plotted in [Figure 16.6](#). The performance of the Van Aerde model is summarized as follows:

- Start-up: The model is able to start the vehicle up from standstill. See [Figure 16.6](#) when $t > 0$ s.
- Speedup: The same argument as in the corresponding part for the Newell nonlinear car-following model applies here. See [Figure 16.6](#) when $0 \text{ s} < t < 100 \text{ s}$.
- Free flow: The model is able to reach and settle at the desired speed under free-flow conditions. See [Figure 16.6](#) when $0 \text{ s} < t < 100 \text{ s}$.
- Cutoff: The same argument as in the corresponding part for the Newell nonlinear car-following model applies here. See [Figure 16.6](#) around $t = 100$ s.
- Following: The model is able to adopt the leader's speed and follow the leader at a reasonable distance. See [Figure 16.6](#) when $100 \text{ s} < t < 200 \text{ s}$.
- Stop and go: The model is able to stop the vehicle safely behind its leader and start the vehicle moving when the leader departs. See [Figure 16.6](#) when $200 \text{ s} \geq t \leq 300 \text{ s}$.
- Trailing: The model is able to speed up normally without being tempted to speed up by its speeding leader. See [Figure 16.6](#) when $300 \text{ s} < t < 400 \text{ s}$.
- Approaching: With limiting deceleration, the model is able to decelerate properly when approaching a stationary vehicle. See [Figure 16.6](#) when $400 \text{ s} \geq t < 420 \text{ s}$.
- Stopping: The model is able to stop the vehicle safely behind the stationary vehicle. See [Figure 16.6](#) when $t \geq 420 \text{ s}$.

The above benchmarking is based on the set of parameters in [Table 16.5](#), and the outcome may differ for a different set of parameters.

Macroscopic Benchmarking

The fundamental diagram implied by the Van Aerde model is presented in [Figure 16.7](#). The model employs four parameters and exhibits a desirable shape with good fitting quality.

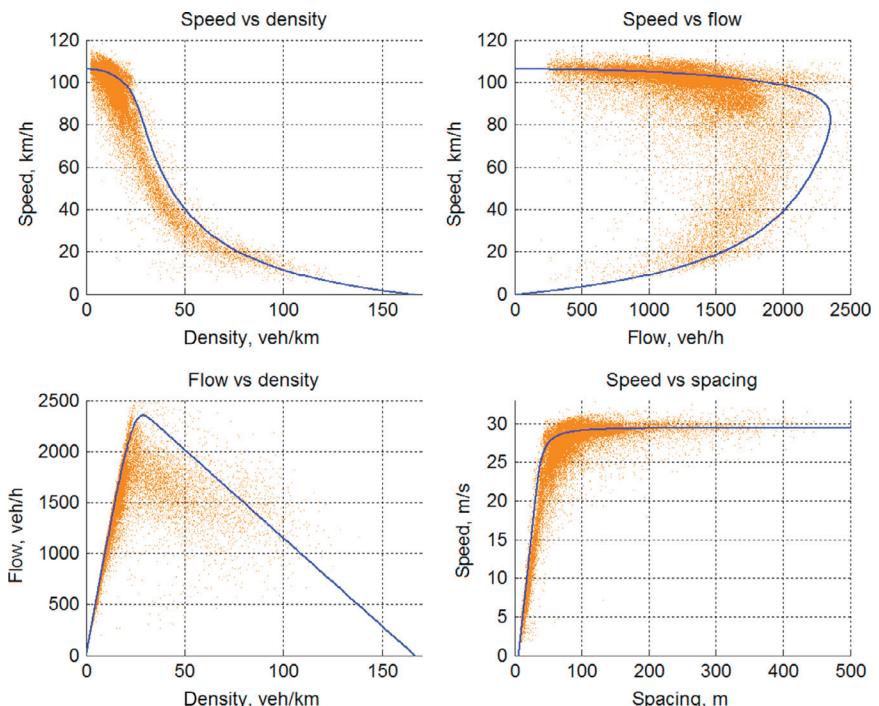


Figure 16.5 Fundamental diagram implied by the IDM.

Table 16.5 Microscopic benchmarking parameters of the Van Aerde model

k_j	v_f	τ_i	v_m	q_m
1/6 vehicles/m	30 m/s	1.0 s	25 m/s	1800 vehicles/h
$x_i(0)$	$\dot{x}_i(0)$	$\ddot{x}_i(0)$		
-99.4 m	0 m/s	0 m/s ²		

Table 16.6 Macroscopic benchmarking parameters of the Van Aerde model

v_f	k_j	v_m	q_m
29.5 m/s	0.25 vehicles/m	20 m/s	1950 vehicles/h

The above benchmarking is based on the set of parameters in Table 16.6, and the outcome may differ for a under different set of parameters.

PROBLEMS

1. Read the capacity condition (q_m, k_m, v_m) off the Newell nonlinear model in Figure 16.2, which is generated with the following parameters: $v_f = 29.5 \text{ m/s}$, $k_j = 0.2 \text{ vehicles per meter}$, and $\lambda = 0.8 \text{ 1/s}$. What capacity condition does the “cloud” (i.e., empirical data) tell you? Comment on how realistic the Newell nonlinear model is when compared with the empirical data.
2. The plot at the bottom right of Figure 16.2 depicts the speed-spacing relationship of the Newell nonlinear model. The curve starts at a point where the spacing is $s = 5 \text{ m}$ and the speed is $v = 0 \text{ m/s}$. Then the curve runs upward with a slope of $\lambda = 0.8 \text{ 1/s}$.
 - a. Assume that this portion of the curve is linear, and establish the underlying linear equation $s = f(v)$.
 - b. Assume a vehicle length of $l = 5 \text{ m}$ and perception-reaction time $\tau = 1.25 \text{ s}$. What is the underlying space-speed relationship—that is, $s = g(v)$ —according to the Forbes model?
 - c. How do you compare the above two models?
 - d. What would you say about the physical meaning of parameter λ ?
3. Show that Van Aerde model is a combination of the Pipes model and the Greenshields model.
4. Vehicle B is following vehicle A according to the Newell simplified car-following model with parameters $\tau = 2 \text{ s}$ and $l = 5 \text{ m}$.

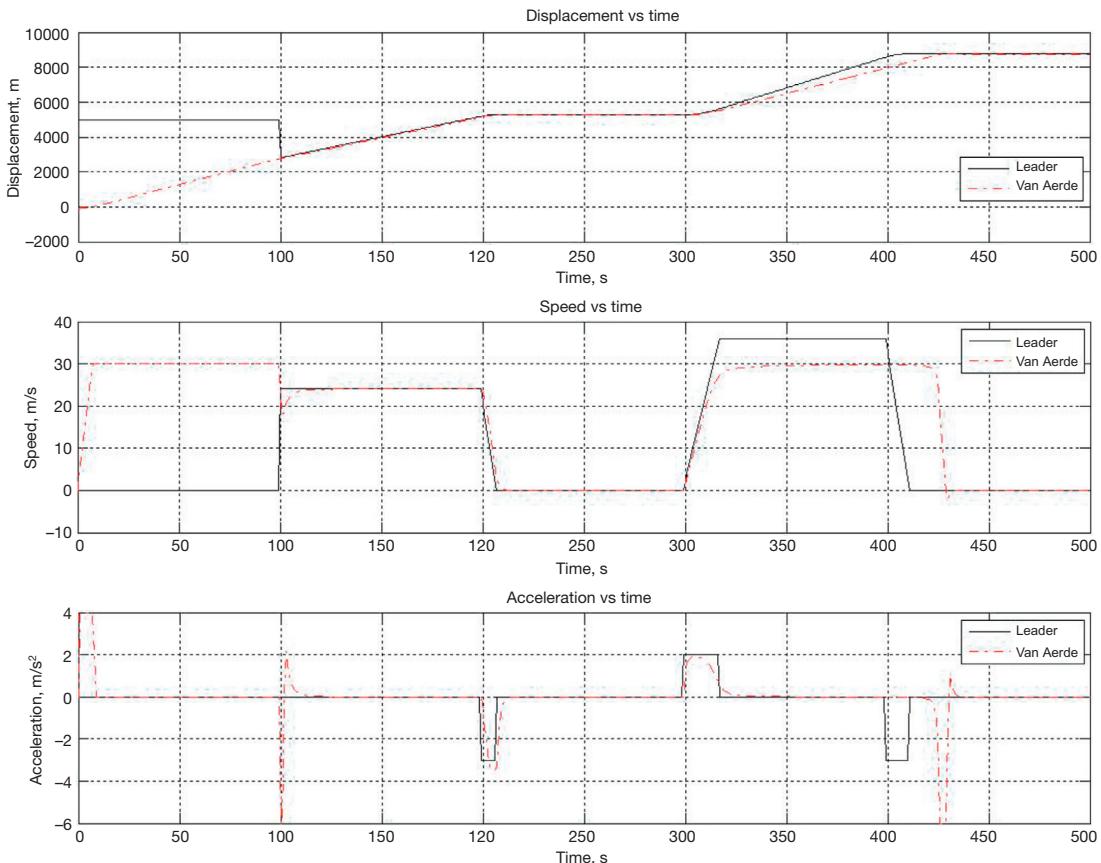


Figure 16.6 Microscopic benchmarking of the Van Aerde model.

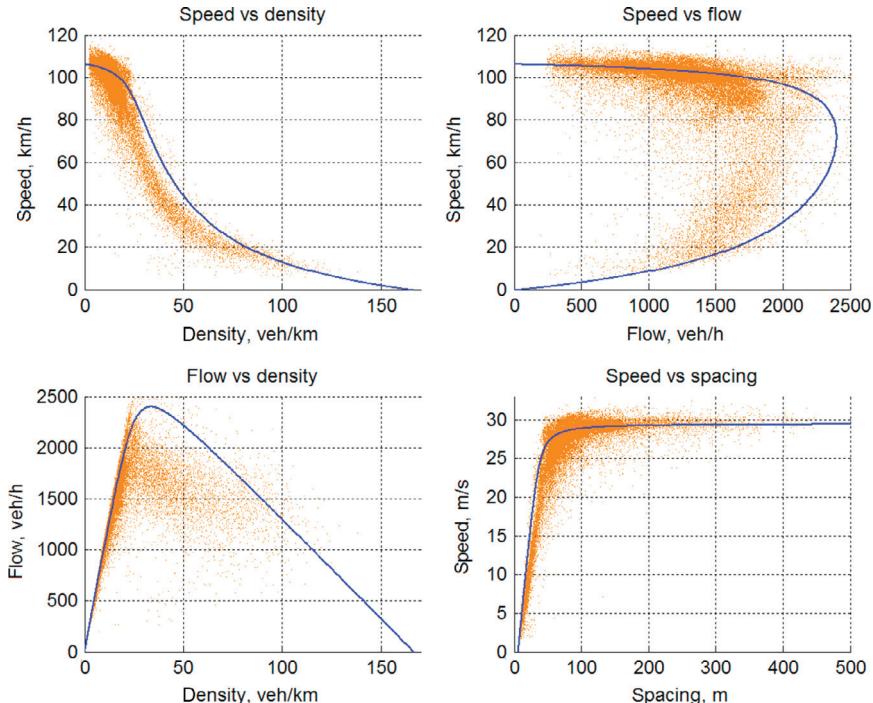


Figure 16.7 Fundamental diagram implied by the Van Aerde model.

- a. Write the underlying car-following model.
 - b. Assume that vehicle A's trajectory is described by the following equation: $x_A(t) = \sqrt{t} - 10$, where $t > 0$. Determine vehicle B's trajectory.
 - c. At time $t = 16$, find the spacing between vehicles A and B.
5. Derive the corresponding density-speed relationship $k = K(v)$ from the IDM under the assumption that $s^* = s_0 + T_i \dot{x}_i$.
6. Further assume $s_0 = 0$ and $\delta = 1$, and derive the corresponding speed-spacing relationship $v = V(s)$ from the above density-speed model.
7. Derive the capacity condition of the Van Aerde model.

CHAPTER 17

More Intelligent Models

The car-following models introduced up to this point share one thing in common: they are one-equation models, except for the Gipps model, which has two equations. This means that these models use a single equation to handle all driving situations, including start-up, speedup, free flow, approaching, following, and stopping. Hence, these models are referred to as single-regime models. The Gipps model is a two-regime model since it has an equation for free flow and another for car following. A model is called a multiregime model if it differentiates driving regimes and handles them using different equations. The car-following models introduced in this chapter fall into this category. In addition, car-following models can mimic the way of human thinking, e.g., using rules and reasoning based on neural networks.

17.1 PSYCHOPHYSICAL MODEL

A typical psychophysical model is the one proposed by Wiedemann [64] in 1974. The model considers two major factors influencing driver's operational control: relative position $\Delta x = x_{i-1} - x_i$ and relative speed $\Delta \dot{x} = \dot{x}_i - \dot{x}_{i-1}$. Hence, the working principle of the model can be illustrated by a diagram with $\Delta \dot{x}$ as the horizontal axis and Δx as the vertical axis (see Figure 17.1).

The operating condition of a vehicle i in relation to its leading vehicle $i - 1$ can be represented as a point $(\Delta \dot{x}, \Delta x)$ in the diagram. As vehicle i moves, its operating point changes accordingly, leaving a trajectory in the diagram. The relation of the two vehicles can be interpreted by examination of the location of the operating point. For example, if the point is on the negative side of $\Delta \dot{x}$, vehicle i is traveling more slowly than vehicle $i - 1$, while the relation is reversed if the point is on the positive side of $\Delta \dot{x}$. In addition, the point is always on the positive side of Δx since vehicle $i - 1$ is in front. The smaller Δx is, the closer the two vehicles are to each other. Hence, the two vehicles collide if Δx is less than one vehicle length l . This

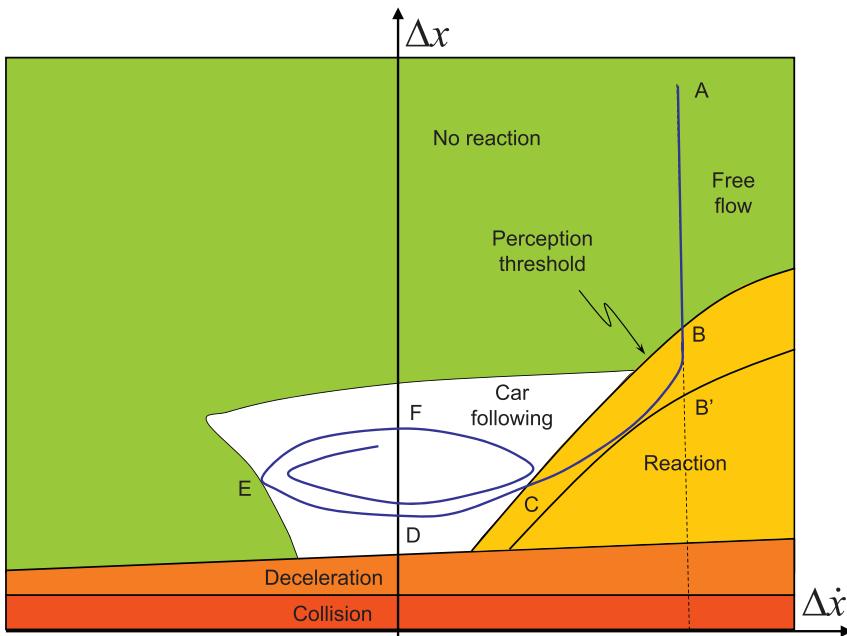


Figure 17.1 Illustration of a psychophysical model.

situation is depicted by the *collision* area in the diagram bounded by the horizontal axis and a horizontal line at $\Delta x = l$.

On top of this area is another area, denoted the *deceleration* area, where the two vehicles are so close that an imminent collision causes the following vehicle to back up for safety.

Now, suppose vehicle i is traveling on a highway with the leading vehicle $i - 1$ far ahead and vehicle i is faster than vehicle $i - 1$. The operating condition can be represented by point A, which has a large positive Δx and a positive $\Delta \dot{x}$. Since vehicle $i - 1$ is far ahead, driver i does not have to respond to vehicle $i - 1$, an area of which is denoted as *no reaction* in the diagram.

As vehicle i keeps moving, the relative speed $\Delta \dot{x}$ remains unchanged, but the relative separation Δx decreases. Hence, the operating point moves downward. Sooner or later, vehicle i will catch up and begin to respond to vehicle $i - 1$ as the gap is closing. However, the cutoff point is rather vague since this is a subjective matter. Perhaps a better way to draw the line is to set an upper limit such as point B, before which drivers are less likely to respond, and a lower limit such as point B', after which drivers definitely need to respond. Note that points B and B' vary as $\Delta \dot{x}$ changes.

The trajectory of point B or point B' under different $\Delta\dot{x}$ separates the *reaction* area from the *no reaction* area.

Since driver i is most likely to respond to vehicle $i - 1$ by slowing down (if lane change is not an option), the operating point moves downward and left toward to point C and finally to point D when the two vehicles are traveling at the same speed. Now the two vehicles are in the car-following regime, during which driver i tries to keep the same pace as vehicle $i - 1$ separated by a comfortable distance. However, drivers are easily bored and distracted, especially during long trips. As a result, driver i might slow down unconsciously (e.g., when using a cell phone). Consequently, $\Delta\dot{x}$ becomes negative and keeps decreasing while Δx increases. As such, the operating point moves from D toward E, at which point the opening gap reminds driver i that he or she is falling behind. Hence, the driver begins to catch up, during which time $\Delta\dot{x}$ increases but is still negative, while Δx keeps increasing. This corresponds to a transition from E to F, when the two vehicles are again traveling at the same speed but with a large gap in between. Next, driver i may want to keep accelerating in order to shorten the gap to a comfortable level, which is denoted as a transition from F back to C. Therefore, as the driver oscillates back and forth around his or her comfortable car-following distance, the operating point drifts around within an area in the diagram denoted as *car following*.

The **psychophysical** model got its name because it involves both psychological activities (such as perception-reaction threshold and unconscious car following) and physical behavior (e.g., accelerating and decelerating efforts). Compared with the models introduced before, this model captures more driving regimes explicitly, such as free flow (*no reaction* area), approaching (*reaction* area), following (*car following* area), and decelerating (*deceleration* area).

17.2 CARSIM MODEL

The CARSIM model [65] is another multiregime model which consists of a set of acceleration algorithms:

A1: Vehicle i is moving but has not yet reached its desired speed v_i .

Depending on vehicle i 's initial speed and the urgency of the task, the acceleration rate is found by from Figures 17.2 and 17.3.

A2: Vehicle i has reached its desired speed v_i . No specific algorithm is provided except that the driver will try to reach v_i as fast as possible while satisfying all safety and operational constraints.

Speed, mph	Cars, ft/s ²	Trucks, ft/s ²	Speed, kph	Cars, m/s ²	Trucks, m/s ²
0-15	8.80	2.20	0-24	2.68	0.67
15-30	5.50	1.10	24-48	1.68	0.34
30-40	5.17	0.88	48-64	1.58	0.27
40-50	4.17	0.44	64-80	1.27	0.13
50-60	3.08	0.44	80-96	0.94	0.13
>60	2.09	0.44	>96	0.64	0.13

Figure 17.2 Typical acceleration rates on a level road.

Speed change, mph	Acceleration, ft/s ²	Deceleration, ft/s ²	Speed change, kph	Acceleration, m/s ²	Deceleration, m/s ²
0-15	4.84	7.77	0-24	1.48	2.37
15-30	4.84	6.74	24-48	1.48	2.05
30-40	4.84	4.84	48-64	1.48	1.48
40-50	3.81	4.84	64-80	1.16	1.48
50-60	2.93	4.84	80-96	0.89	1.48
60-70	1.91	4.84	96-112	0.58	1.48

Figure 17.3 Normal acceleration and deceleration rates for passenger cars.

A3: Vehicle i was stopped and has to start from standstill. A maximum acceleration rate is applied constrained by a noncollision constraint after a response delay.

A4: Vehicle i is in car-following mode with its leader $i - 1$. A4 is determined by the following safety rule being satisfied: vehicle i should leave a nonnegative gap ($s_i - l_{i-1} \geq 0$) from vehicle $i - 1$ should vehicle i be advanced one time step Δt : $s_i(t) = x_{i-1}(t) - x_i(t + \Delta t) \geq l_{i-1}$ where $x_i(t + \Delta t) = x_i(t) + \dot{x}_i \Delta t - 0.5 A_4 \Delta t^2$ and the other variables are as defined before.

A5: Vehicle i in car-following mode is subject to a noncollision constraint which is reinforced by considering the desired spacing:

$$s_i^*(t) = x_{i-1}(t) - x_i(t + \Delta t) \geq \max \left\{ \begin{array}{l} \dot{x}_i(t + \Delta t) \tau_i + l_{i-1} \text{ or} \\ \dot{x}_i(t + \Delta t) \tau_i + \frac{[\dot{x}_i(t + \Delta t)]^2}{2B_i} - \frac{[\dot{x}_{i-1}(t)]^2}{2B_{i-1}} + l_{i-1} \end{array} \right\}$$

where $\dot{x}_i(t + \Delta t) = \dot{x}_i(t) + A_5 \Delta t$, and B_i and B_{i-1} are the maximum deceleration rates of vehicle i and vehicle $i - 1$, respectively. The astute reader immediately recognizes that the first choice of the right-hand side follows the rationale of the Forbes model [53, 54, 66] and the second choice is similar to that of the Gipps model [57] if driver i is willing to apply the emergency brake (i.e., $b_i = B_i$) as well.

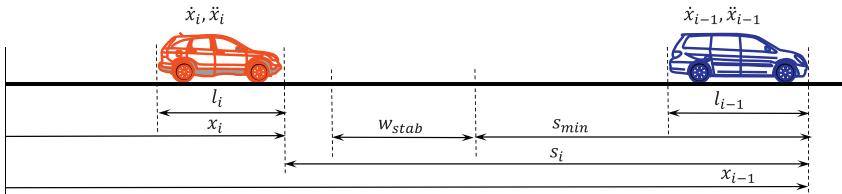


Figure 17.4 Illustration of a rule-based model.

17.3 RULE-BASED MODEL

The model developed by Kosonen [67] is a representative of rule-based models (see Figure 17.4), and is reproduced below:

1. NO SPEED CHANGE

Keep the present speed level (default case).

2. ACCELERATE IF $[\dot{x}_i < v_i]$ and $[t - t_{\text{last}} > T_{\text{acc}}(\dot{x}_i)]$

The current speed \dot{x}_i is less than the desired speed v_i and the time elapsed from the last acceleration t_{last} is more than T_{acc} .

3. NO ACCELERATION IF $[s_{ij} < s_{\min}(\dot{x}_i, \dot{x}_j) + w_{\text{stab}}(\dot{x}_i, \dot{x}_j)]$

The distance from obstacle s_{ij} is less than the minimum safe distance s_{\min} plus the width of the stable area w_{stab} .

4. SLOW DOWN IF $[s_{ij} < s_{\min}(\dot{x}_i, \dot{x}_j)]$

The distance from obstacle s_{ij} is less than the minimum safe distance s_{\min} .

5. DO NOT SLOW DOWN IF $[\dot{x}_i < \dot{x}_j]$ or $[t - t_{\text{last}} < T_{\text{maxdec}}]$

Own speed is less than obstacle speed or maximum deceleration rate is exceeded.

6. GOTO ZERO IF $[s_{ij} < 0]$ and (Obstacle = physical)

Distance to physical obstacle is below zero (= collision).

At each time step, the motion of vehicle i is checked against the above rules one by one. A later rule always supersedes earlier ones should there be a conflict. Compared with the models presented before, the rule-based model is closer to human intelligence with less mathematical tractability.

17.4 NEURAL NETWORK MODEL

Perhaps the approach that best mimics driver behavior is artificial neural networks [68, 69]. This is because artificial neural networks are capable of

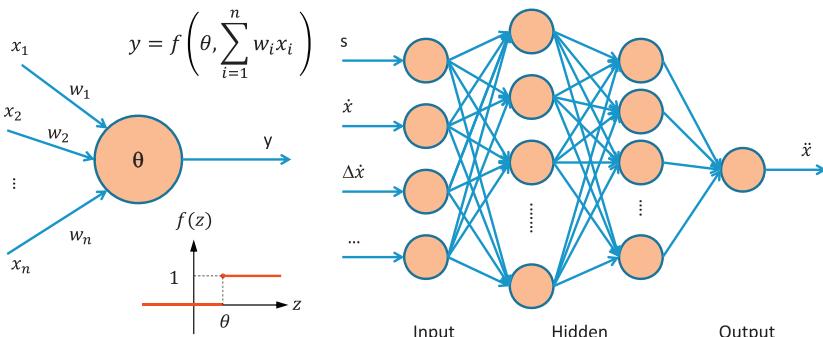


Figure 17.5 Illustration of a neural network model.

associating, recognizing, organizing, memorizing, learning, and adapting. A neural network typically consists of many interconnected working units called neurons; see [Figure 17.5](#) for an example of a neural network in the right panel and a neuron in the left panel. A neuron receives inputs x_1, x_2, \dots, x_n which are weighted w_1, w_2, \dots, w_n , respectively. The total input to the neuron is the weighted sum of individual inputs: $z = \sum_{i=1}^n w_i x_i$. The output of the neuron y depends not only on z but also on the threshold of the neuron θ . The neuron outputs 1 if $z \geq \theta$ and 0 otherwise.

Neurons with such a simple functionality can be organized into neural networks of varying complexity and topology. The right panel in [Figure 17.5](#) illustrates an example of a back-propagation neural network. The network consists of one input layer (which in turn consists of a set of neurons), one output layer, and one or more hidden layers. Each neuron feeds its output only forward to neurons in the next layer, without backward feeding and cross-layer connection.

To apply neural networks to the modeling of car-following behavior, one first identifies a set of factors to be considered that influence the driver's operational control. For example, as discussed before, these influencing factors can be spacing s , speed \dot{x} , relative speed $\Delta\dot{x}$, etc. It is also possible to include other factors not considered before, such as a tailgating vehicle behind, weather, and intervehicular communication. These factors are represented by neurons in the input layer. The output layer in this example consists of only one neuron—acceleration/deceleration or speed choice. If one needs to model not only longitudinal but also lateral motion, a second neuron is necessary to represent steering effort. Between input and output layers lie one or more hidden layers. The more hidden layers the network

has, the more flexible it is, but the more complex it becomes. After the neural network has been constructed, it needs to be trained before it can be useful.

The training process starts with data collection. For example, from field experiments, one observes that, at time t_1 , a vector of input $[s(1), \dot{x}(1), \Delta\dot{x}(1), \dots]$ results in driver operational control $[\ddot{x}(1)]$, and more patterns are observed at t_2, t_3, \dots, t_m :

$$\begin{bmatrix} s(1), \dot{x}(1), \Delta\dot{x}(1), \dots \\ s(2), \dot{x}(2), \Delta\dot{x}(2), \dots \\ \vdots \\ s(m), \dot{x}(m), \Delta\dot{x}(m), \dots \end{bmatrix} \Rightarrow \begin{bmatrix} \ddot{x}(1) \\ \ddot{x}(2) \\ \vdots \\ \ddot{x}(m) \end{bmatrix}. \quad (17.1)$$

After initializing the neural network (i.e., assigning initial values to connection weights and neuron thresholds), one imposes observations at t_1 (i.e., the first row of input data) at the input layer, which feeds forward to hidden layers and eventually to the output layer. If the computed output is different from the observed output, the error is propagated backward layer by layer to adjust their connection weights and neuron thresholds. This is why networks of this kind are called back-propagation networks. After the error has been propagated backward, the same input is imposed again at the input layer and the network computes a new output. This time, the output error, if any, should be smaller than in the previous round. Again, the error needs to be propagated back, and all the weights and thresholds are adjusted for a new round of learning. The process continues until the computed output becomes sufficiently close or equal to the observed output. This completes the learning of the first input-output pattern (i.e., the first row of data set 17.1). Next, one continues with the training of the second row, the third row, and so on. The training is completed after all data in the set have been trained and the neural network is able to associate the correct output with the corresponding input.

The trained neural network is now ready to be applied to vehicle operational control. At any moment, the neural network is able to search for an output (i.e., acceleration/deceleration or speed in the next step) on the basis of the input it receives (i.e., current spacing s , speed \dot{x} , relative speed $\Delta\dot{x}$, etc.). In addition, the neural network may continue learning while working, and hence adapt to a new environment which it has never encountered before.

17.5 SUMMARY OF CAR-FOLLOWING MODELS

It is time to summarize the car-following models introduced so far. One way to classify these models is to look at the model output. *Dynamic models* employ acceleration/deceleration $\ddot{x}_i(t)$ as the model output and the modeling philosophy behind these models is that, at any time, the driver tries to answer the following questions: Should I speed up or slow down next? By how much? Example models for this category are General Motors models (GM models), the IDM, the CARSIM model, the rule-based model, the psychophysical model, and the longitudinal control model, which will be introduced in Chapter 22. *Steady-state models* use speed $\dot{x}_i(t)$ as the model output, and the modeling philosophy behind these models is that, at any time, the driver tries to answer the following question: What is my target speed next? Example models for this category are the Pipes model, the Forbes model, the Gipps model, the Newell nonlinear model, and the Van Aerde model. *Static models* employ displacement $x_i(t)$ as the model output, and the modeling philosophy behind these models is that, at any time, the driver tries to answer the following question: Where should I be next? An example model for this category is the Newell simplified car-following model.

Another way to classify these models is to examine model intelligence. For example, the Pipes model is a one-equation model, and this equation handles all driving situations—that is, they are treated as a single regime. Hence, the Pipes model is a single-regime model. Also in this category are the Forbes model, GM models, Newell models, the IDM, and the Van Aerde model. The Gipps model consists of two equations, one for free flow and the other for car following, and hence is a two-regime model. Both the CARSIM model and the psychophysical model differentiate more driving regimes, and hence are multiregime models. Further, the rule-based model incorporates driving strategies for various driving regimes into a set of simple IF-THEN rules. Better yet, the neural network model applies artificial intelligence to organize, learn, and adapt to driving experiences. Illustrated in [Figure 17.6](#), car-following models become more and more intelligent as one moves from left to right.

On the other hand, since there is only one equation in a single-regime model, it is easy to track the effect of an input on the output. In addition, it is tractable to aggregate/integrate such a microscopic model in order to understand its macroscopic properties. Therefore, single-regime models are mathematically attractive. Two-regime or multiregime models, however,

are inevitably piecewise and involve discontinuity, which makes them less mathematically attractive. Though computationally simple, the rule-based model consists of a set of IF-THEN rules rather than a clearly defined mathematical formulation. Hence, it is very difficult to analyze macroscopic properties of this kind of model. The neural network model, in the extreme, is very intractable because there is no clear mathematical formulation that defines the relation between input variables and the output variable. If a model with a clear mathematical formulation is analogous to a transparent box through which one can trace an input all the way to the output, a neural network is like a black box in which what is happening is a mystery.

A more rigorous effort with regard to the taxonomy of microscopic models was made by the Next-Generation Simulation program¹ (see Figure 17.7). The diagram consists of four modules/rows from top to bottom: *route-choice* models, *lane-changing* models, *gap-acceptance* models, and *car-following* models. In the car-following module, there are a few lines representing different modeling approaches. For example, one approach is called *stimulus-response*, which starts with a few papers published around 1960 serving as the basis of GM models. Labeled along this line are further models that have been proposed or existing models that have been revised, showing the historical evolution of this modeling approach. One line up is the *desired measure* approach, along which are the Pipes model, the Newell nonlinear model, the Gipps model, and the CARSIM model. The next line is the *psychophysical* approach, where one finds the Wiedeman model. This is followed by the *rule-based* approach, an example of which is the Kosonen model. The IDM is on its own at the top. Note that the neural network model a potential addition to this module. The other modules and models shown can be interpreted in a similar way.

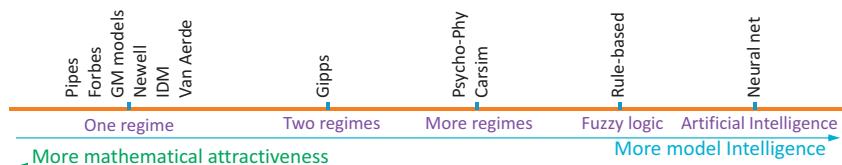


Figure 17.6 Summary of car-following models.

¹ <http://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm>

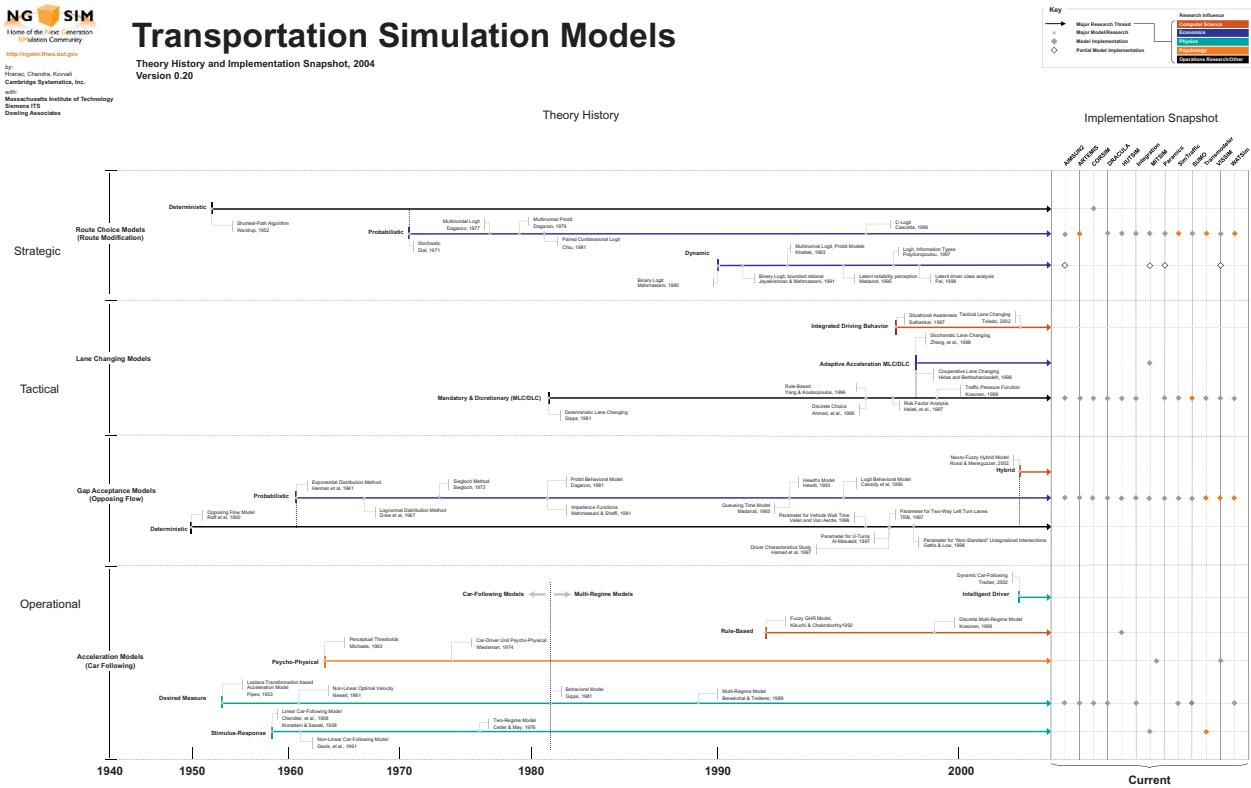
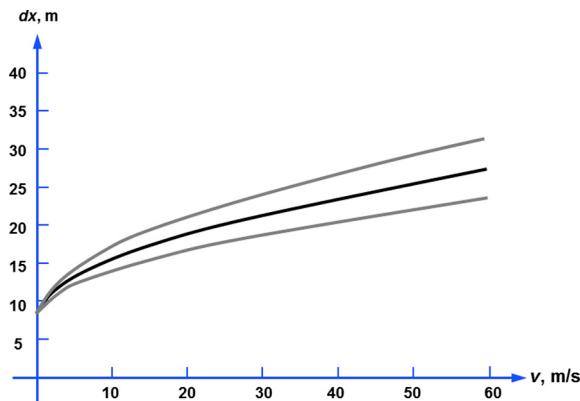


Figure 17.7 Taxonomy of microscopic models.

On the right-hand side of Figure 17.7 there are a set of vertical lines. On top of them are a set of transportation simulators (or simulation software packages), such as AIMSUN, CORSIM, HUTSIM, Integration, Paramics, and VISSIM. The intersection of a horizontal line (a modeling approach) and a vertical line (a simulator) denotes potential implementation of a car-following model of this approach in the simulator. If the implementation is true, a diamond-shaped dot is placed at the intersection. Therefore, it is clear that car following in CORSIM is based on a *desired measure* model, car following in VISSIM is based on a *psychophysical* model, and car following in HUTSIM is based on a *rule-based* model. The connection of simulators and models in other modules can be interpreted in a similar way.

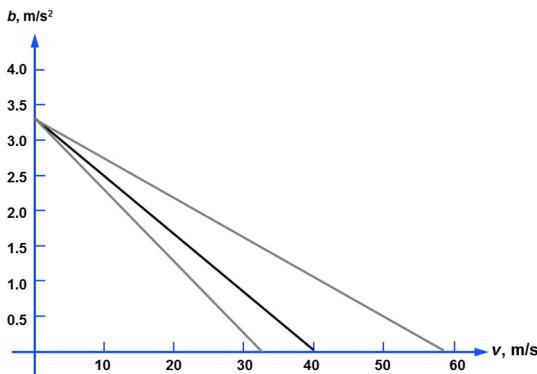
PROBLEMS

- The figure below was used by Wiedemann in his psychophysical model to determine the desired minimum following distance as a function of speed. Use the figure to answer the following questions and do the following task:



- What is the assumed nominal vehicle length (i.e., the average spacing when traffic is jammed)?
 - According to the middle curve, what is the desired minimum following distance when the speed is 30 m/s?
 - Plug the results of (a) and (b) into the Pipes model to estimate the corresponding perception-reaction time.
- The figure below was used by Wiedemann in his psychophysical model to determine the maximum acceleration of passenger cars as a function

of speed. Use the figure to answer the following questions and do the following tasks:



- What is the assumed maximum acceleration?
 - What is the assumed cruise speed according to the middle curve?
 - Formulate the underlying acceleration profile—that is, express acceleration as a function of speed.
 - Derive the corresponding equations to calculate acceleration, speed, and displacement from initial conditions $x(t = 0) = x_0$ and $v(t = 0) = v_0$.
 - If the vehicle is traveling at 20 m/s, determine its speed after 5 s of acceleration and the distance traveled during that time.
3. Figure 17.2 depicts typical acceleration rates on a level road used by the CARSIM model. Convert the units to the metric system and plot the data on top of the figure in problem 2 (use the column for cars and take the midpoint of each speed range). Comment on how the acceleration profiles differ in the CARsim model and the psychophysical model.
4. Determine all the rules that apply and the rule that actually takes effect in each of the following scenarios according to the rule-based model:
- A vehicle is entering an empty freeway at a speed of 60 km/h. The freeway speed limit is 90 km/h.
 - A vehicle is cruising on the freeway at a desired speed of 95 km/h. There is no other vehicle in the visible range in front.
 - A vehicle at a speed of 100 km/h is approaching a leader at a speed of 90 km/h. Their current spacing is 70 m, and the minimum safe distance 100 m.

- d. A vehicle with a desired speed of 95 km/h is following its leader, both traveling at a speed 90 km/h with a spacing of 120 m. The minimum safe distance is 100 m, and the stable area is 25 m.
 - e. A vehicle at a speed of 70 km/h is changing to the target lane, where there is a leader traveling at a speed of 90 km/h. The spacing between the two vehicles is 70 m, and the minimum safe distance is 100 m.
5. The minimum safe distance in the rule-based model can be formulated in many ways. Use your car-following knowledge learned in previous chapters to propose two ways to determine the minimum safe distance.

CHAPTER 18

Picoscopic Modeling

Suppose one is observing traffic 10,000 m above the ground, and the traffic behaves as a compressible fluid whose states (speed, flow, density, etc.) propagate back and forth like waves. This is a scenario of *macroscopic* modeling. If one goes to 3000 m above the ground, the sense of waves recedes and a scene of particles emerges. A vehicle behaves as a particle hopping from one cell to another governed by predetermined logic. This is a scenario of *mesoscopic* modeling. If one goes even lower to 1000 m above the ground, the scene is dominated by moving particles which interact with each other so as to maintain safe positions in the traffic stream. This is a scenario of *microscopic* modeling as well as the state of the art.

What is the next level of traffic flow modeling? Continuing with the above analogy, the next level should provide a perspective as if one were on the ground and driving in one of the vehicles in the traffic. What one sees now is neither a wave nor a particle, but a detailed picture incorporating drivers, vehicles, and the environment (e.g., roadway, signs, signals) (see Figure 18.1). Drivers collect information and make control decisions in terms of steering, acceleration, and deceleration. Vehicles dynamically respond to their drivers by executing their control decisions and moving on the ground accordingly. Feedback from vehicle dynamics, together with information from the environment, constitutes the basis for drivers to make control decisions in the next step. Traffic operation is simply the movement and interaction of all vehicles in the system over time and space. This is a scenario of *picoscopic* modeling.

18.1 DRIVER, VEHICLE, AND ENVIRONMENT

Traffic flow modeling at the picoscopic level should not only represent drivers, vehicles, and the environment in different models, but should also capture the interaction among these components. Therefore, a natural approach is to address the modeling problem as a driver-vehicle-environment closed-loop system [70, 71] as illustrated in Figure 18.2.

In a transportation system, drivers are active components which make decisions, while vehicles are passive components which execute decisions.



Figure 18.1 A picoscopic view of a transportation system.

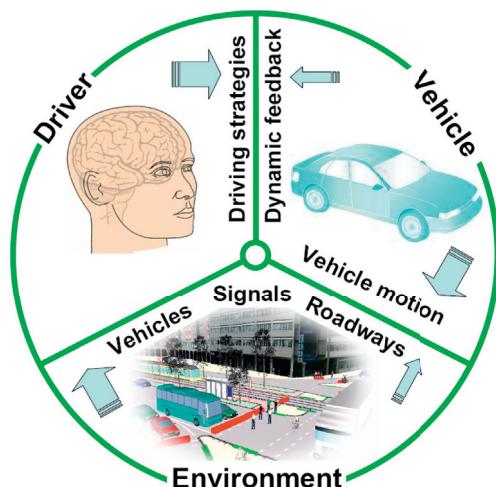


Figure 18.2 A driver-vehicle-environment closed-loop system.

The interaction between a driver and his/her vehicle constitutes a basic unit in a traffic stream. Therefore, a natural way to mimic the real-world system is to model drivers and vehicles separately but with interaction between them. Drivers are motivated by goals, act autonomously, and reason on the basis of their knowledge. [Figure 18.3](#) presents the structure of such a driver modeling approach.

This approach involves three components: inputs, driver, and outputs. Inputs to the model are environment information and vehicle feedback. The environment loosely refers to the entire system, including drivers, vehicles, pedestrians, roadway infrastructure, traffic control devices, roadside, abutting lands, nearby business, etc. Vehicle feedback includes part of vehicle dynamic responses, such as vehicle speed, acceleration, and yaw

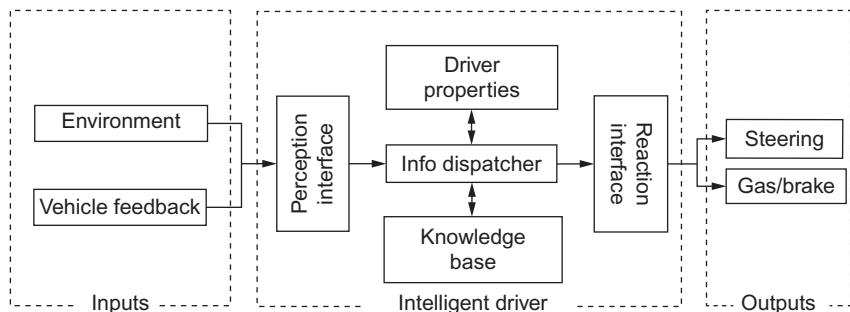


Figure 18.3 Picoscopic modeling: driver modeling.

velocity, perceived by the driver and affecting his/her driving decision. As an intelligent agent, a driver is able to (a) respond in a timely fashion to changes in the environment, (b) exercise control over his/her own actions, (c) pursue a goal which motivates his/her actions, (d) communicate with other agents, and (e) change his/her behavior on the basis of previous experience. With these considerations, the driver model consists of the following components: a perception interface which collects and transforms information before it enters the driver, a reaction interface which converts driver decisions to actionable instructions before they are executed by the vehicle, driver properties including driver's goals and characteristics, a knowledge base including experiences and decision rules that govern driving behavior, and an information dispatcher, which is the central processing unit of the driver. Outputs of the driver model are driving decisions, including steering, accelerating, and braking.

In Chapter 21, a field theory will be introduced that can serve as the basis for the intelligent driver. In this theory, highways and vehicles are perceived as a field by a subject driver whose driving strategy is to navigate through the field along its valley.

The approach to vehicle modeling needs to incorporate vehicle dynamics so that vehicle dynamic responses and lateral movement can be captured. [Figure 18.4](#) illustrates such an approach which includes inputs, dynamic vehicle, and outputs. Inputs to the vehicle come from two sources: inputs from the driver, including steering, throttle position, and brake position, and inputs from environment such as roadway surfaces, lanes, curves, and resistances. The vehicle model consists of vehicle-specific information (i.e., vehicle properties such as mass, dimension, and engine power) and vehicle-generic information, including a set of dynamic equations describing the

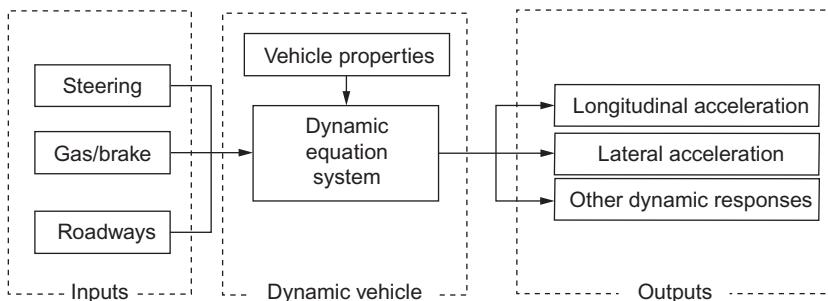


Figure 18.4 Picoscopic modeling: vehicle modeling.

dynamic performance of a class of vehicles, such as acceleration/deceleration and steering performance. Outputs of the dynamic vehicle are vehicle dynamic responses, of which longitudinal acceleration, lateral acceleration, and yaw velocity are of particular interest.

In Chapter 19, a simple engine model will be formulated with reasonable accuracy and excellent computational efficiency to facilitate vehicle modeling. In Chapter 20, a simple dynamic interactive vehicle model will be formulated that requires minimal calibration effort and computational resources.

Combining the above driver and vehicle models results in a driver-vehicle unit which constitutes a basic building block of roadway traffic. Such units as well as roadways, traffic control devices, and other transportation system components constitute a general environment in which a driver-vehicle unit operates. The interactions among drivers, vehicles, and the environment are summarized in the picoscopic transportation modeling architecture shown in [Figure 18.5](#).

In this architecture, the driver receives information from the environment such as roadways, traffic control devices, and the presence of other vehicles. The driver also receives information from his/her own vehicle such as speed, acceleration, and yaw velocity. These sources of information, together with driver properties (such as characteristics and goals), are used to determine driving strategies (such as steering and accelerating/braking). The driving strategies are fed forward to the vehicle, which also receives roadway information from the environment. These sources of information, together with vehicle properties, determine the vehicle's dynamic responses on the basis of vehicle dynamic equations. Moving longitudinally and laterally, the vehicle constitutes part of the environment. Some of vehicle dynamic responses such as speed, acceleration, and yaw velocity are fed back

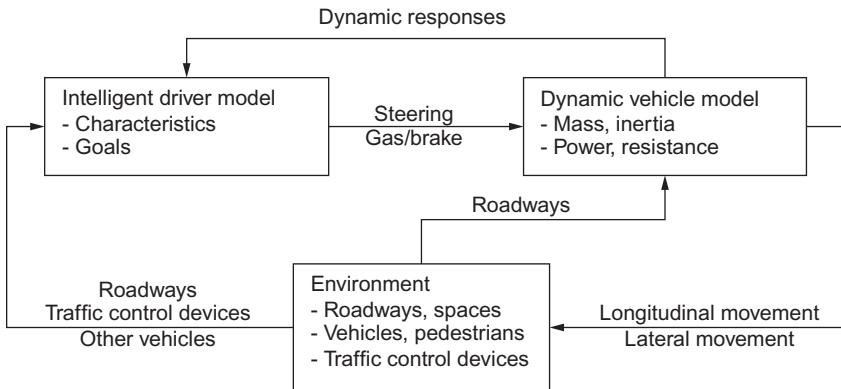


Figure 18.5 Picoscopic modeling: modeling architecture.

to the driver to determine driving strategies in the next step. Therefore, the architecture creates an environment, in which each vehicle is an autonomous agent which is driven by goals and is able to achieve the goals by moving through the environment. Thus, traffic operation is simply the motion and interaction of all vehicles in the environment.

18.2 APPLICATIONS OF PICOSCOPIC MODELING

Transportation modeling and simulation is characterized by two competing dimensions: scale (i.e., geographical scope covered in the modeling) and level of detail (i.e., resolution provided by the model). Because of the processing power of today's computers, a macroscopic model can achieve a very large modeling scale, such as the Commonwealth of Massachusetts, with relatively low resolution. A mesoscopic model strikes a balance between the two; a microscopic model is able to provide fine modeling resolution within a limited geographical area, such as the city of Boston. Following this trend, a picoscopic model would furnish ultrahigh modeling resolution but within a very limited geographical area, such as the roads surrounding Public Garden in Boston. With such a fine level of detail, picoscopic modeling can help address many transportation-related problems, among which the following are a few examples.

18.2.1 Interactive Highway Safety Design

Picoscopic transportation modeling can be used to assist highway design. For example, a highway design can be tested by different “drivers” and

“vehicles” in a computer to check if the highway provides sufficient sight distance to avoid accidents or a curve is properly superelevated to allow safe turning. Such an interactive highway safety design not only ensures design quality but also saves time and resources to achieve the design goal.

18.2.2 Connected Vehicle Technology

Future vehicles will be equipped with dedicated short-range communications, along with sensing, positioning, and computing devices. As such, vehicles will be able to communicate with other vehicles as well as the roadside. Such a connected vehicle technology will transform future highways and streets into an environment that encompasses ubiquitous computing and communication (see Figure 18.6). Consequently, innovative applications can be deployed to dramatically increase safety, throughput, and energy efficiency. However, such systems elude mathematical analysis and conventional simulation because of the complexity and interdependency involved. Picoscopic modeling might be able to address these systems because it not only captures sufficient modeling details but also allows the incorporation of the effects of connected vehicle technology into modeling.

18.2.3 Transportation Forensics

Investigation of a traffic accident frequently requires the ability to decipher what happens shortly before, during, and after the accident. This involves reconstruction of the accident during which the driver perceives

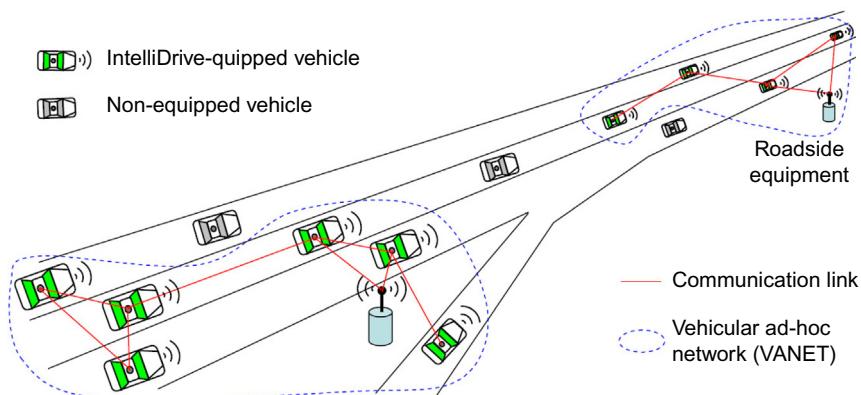


Figure 18.6 An illustration of connected vehicle technology.

an immediate hazard, makes a decision, and executes control, while the vehicle's dynamically responds to control instructions, moves on the ground, collides with another vehicle, and is redirected, potentially causing a second crash. Modeling at such a level of detail necessitates a picoscopic approach.

18.2.4 Emergency Management

In analyzing transportation systems under extreme conditions, one must have both the capability of overseeing the full picture (e.g., a regional transportation network) and the capability of zooming in for local details (e.g., a corridor or an intersection). Anyone who is familiar with Google Maps or Google Earth develops a sense of the importance of having global information yet being able to zoom in and view local details. The transportation modeling and simulation tools developed so far have offered only a single-level resolution. As such, they are suited for either solving large-scale transportation problems with coarse resolution or solving small-scale problems with fine details. Though these tools can provide a partial solution, efforts are needed to integrate them to provide an integral analysis with both scale and detail because emergency management involves addressing multiple aspects of the emergency.

Transportation modeling at the picoscopic level is essential to help achieve very fine modeling detail and address problems that are beyond the capabilities of conventional modeling tools. For example, conventional tools are developed for use under peaceful settings, and thus they are not suited for coping with unusual traffic operations. Under extreme conditions, drivers are under great pressure, and their driving behavior changes drastically from their usual ways. As a result, safety as a primary goal may give way to the need of getting out of the endangered site as quickly as possible. Traffic rules may not be observed, and consequently, unusual operations such as running a red light, violating priority rules, and off-road operations are possible. Existing modeling and simulation tools are based on the assumption of driving in a safe world, so they have difficulty to replicate situations under extreme conditions. Moreover, panic behavior is likely to result in more frequent accidents and crashes than usual. However, existing modeling and simulation tools are designed to guarantee "accident-free" situations, which prevents these tools from modeling and simulating transportation systems under extreme conditions.

PROBLEMS

1. Where do car-following models fit in the driver-vehicle-environment closed-loop system framework?
2. Comment on the potential benefits of picoscopic traffic flow modeling.
3. Comment on the potential costs of picoscopic traffic flow modeling.
4. Comment on the appropriate applications of picoscopic traffic flow modeling.

CHAPTER 19

Engine Modeling

As discussed in the previous chapter, a key component of picoscopic modeling is to capture vehicle dynamics. Such an issue has been greatly simplified in microscopic modeling where the driver and vehicle are combined into a single unit called an active particle. Vehicle dynamic properties are either ignored or simplified. For example, in steady-state car-following models such as the Pipes model, a vehicle's speed can jump from zero to an arbitrary full speed, resulting in an unrealistic acceleration profile. In dynamic car-following models such as General Motors models, the rate of acceleration can change in such a way that is beyond the capability of real-world vehicles. However, in picoscopic modeling, finer details and accuracy are called for, and hence the above simplified approach becomes inadequate. Because vehicle acceleration plays an important role in traffic operation such as in the calculation safe car-following distances, the determination of acceptable gap sizes, and bypassing slow vehicles, it is critical to base the modeling on sound dynamic vehicle models. Toward this goal, engine modeling is the first step which determines the available power and further torque under varying speeds. Then, the power and torque are fed into dynamic vehicle models to determine the vehicle's acceleration capability.¹

19.1 INTRODUCTION

Though there has been a wealth of literature in the modeling of internal combustion (IC) engines, these models were developed with a special interest in assisting engine design, analysis, control, and diagnosis. While these models are quite successful for their intended purposes, several reasons prevent them from being equally successful in traffic flow modeling and further the modeling of connected vehicle technology (CVT). For example, a typical procedure in these applications is to invoke routines such as car-following, lane-changing, and gap-acceptance logic to check for potential collisions. To ensure safety, this procedure has to be repeated with such a high frequency that conventional engine models, owing to their intrinsic complexity, are beyond the capacity of a contemporary onboard computer.

¹ This chapter is reproduced from [72].

In addition, most of these engine models require proprietary parameters such as throttle body size and mass of the piston. This prevents the adoption of these models across a wide variety of vehicles. Therefore, an ideal engine model suited for the above-mentioned applications should meet the following criteria:

- Accuracy: The engine model must provide reasonable accuracy to predict engine performance with throttle and engine speed as inputs and engine power and torque as outputs.
- Computational efficiency: The engine model must be simple enough to facilitate onboard computing with high frequency in real time.
- Accessibility: To assist wide deployment across different vehicles, the engine model should not rely on proprietary parameters and variables that are difficult to obtain. All the information needed to run the model (such as peak engine power, torque, and the associated engine speeds) should be publicly available (e.g., <http://www.cars.com>).
- Formulation: The engine model should be analytical. Engine models based on lookup tables are not only prohibitive to prepare wide classes of vehicles but are also resource-demanding in computation and storage.
- Calibration: The engine model should involve the least calibration effort or better yet should be calibration-free. Again, it would be a daunting task if an engine model had to be calibrated for every vehicle.

With the above list of criteria, the objective of this chapter is to develop a simple engine model that is suited for these applications. Three simple engine models are presented in this chapter. These engine models will be formulated and empirically validated. Special attention will be paid to the above criteria when we compare the performance of these models, on the basis of which the best model will be recommended. Compared with existing work reviewed in the next section, a limited theoretical contribution is claimed in that these models are rather simple and some of the modeling concepts (such as polynomial fitting and the Bernoulli principle) have already been explored in the past. However, the recommended model does fill a gap in a nonconventional arena such as CVT-enabled applications, where excellent computational efficiency and reasonable accuracy are desirable.

19.2 REVIEW OF EXISTING ENGINE MODELS

The objective of this section is to highlight existing work in engine modeling with an emphasis on IC engines. Given the wealth of literature, it

is practically intractable to include all work. Nevertheless, the review should present a reasonable overview of historical efforts and the current state.

It appears that efforts to model engines have a much shorter history than the engine itself. The first physically based dynamic engine models were reported in Refs. [73–76] which recognized the effects of throttle and intake manifold dynamics. Much of the early effort in engine modeling was surveyed in Ref. [77], with a focus on IC engine models for control.

A trend of increasing modeling accuracy was quite noticeable in the historical evolution. The engine models developed in Refs. [78, 79] included fuel film dynamics and engine rotational dynamics with transport delays. Continuing this modeling approach, a three-state engine model was developed in Ref. [80] based on the work in Refs. [74, 75, 81]. Shortly afterward, Akinci et al. [82] also presented a nonlinear three-state dynamic model of a spark-ignition engine and further effort was reported in Ref. [83]. Rizzoni [84] formulated a global model for the IC engine, and a concurrent paper [85] described a stochastic model. A nonlinear engine model was proposed in Ref. [86]. Hong [87] developed an engine model based on the “filling and emptying” method for unsteady gas flow across the engine cylinder [88]. A low-dimensional, physically motivated engine model was proposed in Ref. [89]. Shiao et al. [90] proposed remedies to the assumption of constant mass moments of inertia which had led many engine models to perform poorly under high engine speed. To serve the purpose of engine design, Chiavola [91] described the unsteady gas flow in both intake and exhaust systems. A very complicated engine model involving 12 degrees of freedom [92] was proposed to capture even more details.

On the applied side, efforts were identified which adopted existing models or extended existing work. Kabganian and Kazemi [93] applied the two-state engine model developed in Ref. [80] to slip control. A real-time engine model [94] similar to that in Ref. [78] was used to develop a nonlinear model-based control strategy for hybrid vehicles. Delprat et al. [95] modeled an IC engine as part of hybrid vehicle modeling. Scillieri et al. [96] developed a direct-injection spark-ignition engine model to demonstrate the potential performance benefits of reference feed-forward control. Two simulation packages involving IC engine models [97, 98] were also identified.

In contrast to the ever-increasing desire for modeling details, some applications such as real-time engine control necessitate simpler engine models. Recognizing the inherently nonlinear nature of IC engines, Cook and Powell [99] argued that a linear engine model reduced from the model

in Ref. [73] might be desirable for the purpose of engine control analysis. To facilitate the development of autonomous intelligent cruise control), Swaroop et al. [100] used an engine model which was essentially the first state equation developed in Ref. [80]. A very simple engine model was presented in Ref. [101] for teaching purposes. An even simpler model was suggested by Genta [102] to assist the modeling of vehicle dynamics, and we shall revisit this model shortly.

To facilitate a cross-comparison of engine models in terms of their complexity, accuracy, accessibility, and intended applications, a summary table is provided in [Appendix 19.A](#).

19.3 SIMPLE MATHEMATICAL ENGINE MODELS

This section presents three simple engine models. Model I is an existing model [102], while models II and III were developed by the author.

19.3.1 Model I: Polynomial Model

In an effort to develop a dynamic vehicle model, Genta [102] suggested a very simple engine model which used a polynomial to empirically approximate the relationship between engine power, P , and engine speed, ω —that is,

$$P = \sum_{i=1}^3 C_i \omega^i, \quad (19.1)$$

where the C_i ($i = 0, 1, 2, 3$) are coefficients and can be estimated from empirical engine curves. Artamonov et al. [103] suggested the following values for a spark-injection engine:

$$\begin{aligned} C_1 &= P_{\max}/\omega_p, \\ C_2 &= P_{\max}/\omega_p^2, \\ C_3 &= -P_{\max}/\omega_p^3, \end{aligned} \quad (19.2)$$

where P_{\max} is the peak power and ω_{\max} is the engine speed at which the power peaks. As is well known, engine torque, Γ , is engine power divided by engine speed:

$$\Gamma = \sum_{i=1}^3 C_i \omega^{i-1}, \quad (19.3)$$

where coefficients C_i ($i = 0, 1, 2, 3$) remain the same as in Equation 19.1.

19.3.2 Model II: Parabolic Model

Motivated by the simplicity of model I and noticing the peak in a typical engine torque curve, one conjectures that a parabola might suffice to approximate the torque curve:

$$\Gamma = C_1 + C_2(\omega - \omega_t)^2, \quad (19.4)$$

where C_1 and C_2 are constants and ω_t is the engine speed at peak torque. To ensure that the power curve peaks at ω_p , one replaces C_1 with a different coefficient C_3 :

$$P = C_3\omega + C_2(\omega - \omega_t)^2\omega. \quad (19.5)$$

Given that the engine outputs P_{\max} at ω_p and outputs Γ_{\max} at ω_t , the following result:

$$\Gamma_{\max} = C_1 + C_2(\omega_t - \omega_t)^2 = C_1, \quad (19.6)$$

$$P_{\max} = C_3\omega_p + C_2(\omega_p - \omega_t)^2\omega_p, \quad (19.7)$$

$$\frac{dP}{d\omega}|_{\omega=\omega_p} = (C_3 + C_2(\omega - \omega_t)^2 + 2C_2\omega(\omega - \omega_t))|_{\omega=\omega_p} = 0. \quad (19.8)$$

Solve Equations 19.7 and 19.8:

$$C_2 = -\frac{P_{\max}}{2\omega_p^2(\omega_p - \omega_t)}, \quad (19.9)$$

$$C_3 = \frac{P_{\max}}{2\omega_p^2}(3\omega_p - \omega_t). \quad (19.10)$$

Therefore,

$$\Gamma = \Gamma_{\max} - \frac{P_{\max}}{2\omega_p^2(\omega_p - \omega_t)}(\omega - \omega_t)^2, \quad (19.11)$$

$$P = \frac{P_{\max}}{2\omega_p^2}(3\omega_p - \omega_t)\omega - \frac{P_{\max}}{2\omega_p^2(\omega_p - \omega_t)}(\omega - \omega_t)^2\omega. \quad (19.12)$$

Equations 19.11 and 19.12 constitute model II, and guarantees that its power and torque curves peak at their respective peak engine speeds.

19.3.3 Model III: Bernoulli Model

This model is based on the Bernoulli principle, which states that for an ideal fluid (e.g., air) on which no external work is performed, an increase in velocity occurs simultaneously with a decrease in pressure or a change in the fluid's gravitational potential energy. When the fluid flows through a pipe

(e.g., the intake manifold) with a constriction (e.g., the throttle) in it, the fluid velocity at the constriction must increase in order to satisfy the equation of continuity, while its pressure must decrease because of conservation of energy. The limiting condition of this effect is choked flow, where the mass flow rate is independent of the downstream pressure (e.g., in the combustion chamber), depending only on the temperature and pressure on the upstream side of the constriction (e.g., the atmosphere). The physical point at which the choking occurs is when the fluid velocity at the constriction is at sonic conditions or at a Mach number (the ratio of fluid velocity and sound speed) of 1. With the above knowledge, the Bernoulli engine model is developed as follows.

The theoretical volumetric fresh mixture flow rate into the engine, \dot{V}_t , is

$$\begin{aligned}\dot{V}_t \text{ (m}^3/\text{s)} &= V_e \text{ (m}^3/\text{cycle)} \times \text{cycles/revolution} \\ &\quad \times \text{enginespeed(revolutions/s)},\end{aligned}\quad (19.13)$$

where V_e is engine displacement, the number of cycles per revolution is $1/2$ for a four-stroke engine, and the engine speed (revolutions per second) is $\omega_e/2$, where ω_e is the engine speed in radians per second. Therefore,

$$\dot{V}_t = V_e \times \frac{1}{2} \times \frac{\omega}{2\pi} = \frac{V_e \omega_e}{4\pi}. \quad (19.14)$$

This model assumes that the air is an ideal gas. According to the ideal gas law,

$$pV = \frac{m}{m'}RT, \quad (19.15)$$

where p is the absolute pressure, V is the volume of the vessel containing the gas, m is the mass of the gas, m' is the molar mass of the gas, R is the gas constant, and T is the temperature in kelvins. Therefore, $m = \frac{pm'V}{RT}$, and the density of the gas in the vessel is

$$\rho = \frac{m}{V} = \frac{pm'}{RT} = \frac{p}{R_a T}, \quad (19.16)$$

where $R_a = R/m'$ and for air $R_a \approx 287 \text{ N m/kg/K}$. Further, the mass air flow rate, \dot{m} , as a function of the volumetric air flow rate, \dot{V} , is

$$\dot{m} = \frac{pm'}{RT} \dot{V} = \frac{p}{R_a T} \dot{V}. \quad (19.17)$$

For an engine, \dot{V} is replaced by \dot{V}_t and the speed of air flow is $v = \dot{V}/A$, where A is the cross-sectional area of any point in the intake manifold. The

constriction in the manifold is the throttle, whose cross-sectional area is $\theta \times A$, where θ is percent of throttle opening. So the mass flow rate of air entering the engine is

$$\dot{m} = \frac{p}{R_a T} \dot{V}_t = \frac{p}{R_a T} \nu A. \quad (19.18)$$

According to compressible fluid mechanics [104], the speed of air flow, ν , is related to a Mach number, M_a , which is the ratio of air flow speed to sound speed $\nu_s = \sqrt{kR_a T}$ —that is,

$$M_a = \frac{\nu}{\nu_s} = \frac{\nu}{\sqrt{kR_a T}} = \frac{\dot{V}_t}{A\sqrt{kR_a T}}, \quad (19.19)$$

where k is the specific heat ratio. Assume the stagnation state (where the flow is brought into a complete motionless condition in an isentropic process without other forces) holds. With the stagnation state for the ideal gas model in Sections 4.1 and 4.2 in Ref. [104], Equation 19.18 can be translated to

$$\dot{m} = A \left(\frac{\sqrt{k} M_a p_0}{\sqrt{R_a} T_0} \right) \left(1 + \frac{k-1}{2} M_a^2 \right)^{-\frac{k+1}{2(k-1)}}, \quad (19.20)$$

where p_0 and T_0 are the stagnation pressure and temperature, respectively. Plugging 19.19 into 19.20 yields

$$\dot{m} = A \left(\frac{\dot{V}_t p_0}{A R_a T_0} \right) \left(1 + \frac{\dot{V}_t^2 (k-1)}{2 A^2 k R_a T_0} \right)^{-\frac{k+1}{2(k-1)}}. \quad (19.21)$$

Notice that Equations 19.20 and 19.21 apply to flow everywhere. When the flow is choked (i.e., $M_a = 1$) and the stagnation conditions (i.e., temperature, pressure) do not change, Equation 19.20 reduces to

$$\dot{m} = A \left(\frac{\sqrt{k} p_0}{\sqrt{R_a} T_0} \right) \left(1 + \frac{k-1}{2} \right)^{-\frac{k+1}{2(k-1)}}. \quad (19.22)$$

For exact stoichiometric air-fuel ratio λ , fuel energy density E_f , and engine thermal efficiency η , the power developed by the engine is

$$P = \lambda E_f \eta \left[A \left(\frac{\dot{V}_t p_0}{A R_a T_0} \right) \left(1 + \frac{\dot{V}_t^2 (k-1)}{2 A^2 k R_a T_0} \right)^{-\frac{k+1}{2(k-1)}} \right]. \quad (19.23)$$

Plugging in [equation 19.14](#), we obtain

$$P = \lambda E_f \eta \left[A \left(\frac{V_e \omega_e p_0}{4\pi A R_a T_0} \right) \left(1 + \frac{V_e^2 \omega_e^2 (k-1)}{32\pi^2 A^2 k R_a T_0} \right)^{-\frac{k+1}{2(k-1)}} \right]. \quad (19.24)$$

The torque that the engine develops is

$$\Gamma = \lambda E_f \eta \left[A \left(\frac{V_e p_0}{4\pi A R_a T_0} \right) \left(1 + \frac{V_e^2 \omega_e^2 (k-1)}{32\pi^2 A^2 k R_a T_0} \right)^{-\frac{k+1}{2(k-1)}} \right]. \quad (19.25)$$

Empirical comparison shows that this model explains engine performance quite well up to peak torque and power. However, there are considerable differences between the model and the empirical engine curves after peak torque and power. Therefore, the engine model is modified by the addition of a correction term:

$$P = \lambda E_f \eta \left[A \left(\frac{V_e \omega_e p_0}{4\pi A R_a T_0} \right) \left(1 + \frac{V_e^2 \omega_e^2 (k-1)}{32\pi^2 A^2 k R_a T_0} \right)^{-\frac{k+1}{2(k-1)}} \right] - \alpha P_{\max} e^{\frac{\beta(\omega - \omega_p)}{\omega_p}}, \quad (19.26)$$

where α and β are coefficients to be calibrated. The specific form of the correction term is obtained mainly by trial and error from fitting a wide variety of engine power curves. This model, because of its simplicity, captures only the major aspect of an engine. Since many of the engine details are left out, the model exhibits only moderate accuracy even with the correction term. We also recognize that the concept of this Bernoulli principle-based model is not new, and a similar discussion can be found in existing work, such as [105].

19.4 VALIDATION AND COMPARISON OF THE ENGINE MODELS

To validate the three engine models as well as to compare their relative performance, we need empirical engine power and torque curves. Unfortunately, we do not have much choice because such empirical data are typically proprietary unless they are made available by interested parties. Provided in this validation study are empirical curves for the following four automotive engines: 2008 Mercedes CLS, 2006 Honda Civic, 2006 Pagani Zonda, and 1964 Chevrolet Corvair. Hopefully, these engines provide a good representation of vehicle makes, models, and model years. Technical specifications of these engines are listed in [Table 19.1](#). Additional

Table 19.1 Technical specifications of engines used in the validation study

Engine Tech specification	Mercedes CLS 2008	Honda Civic 2006	Pagani Zonda 2006	Chevrolet Corvair 1964
Peak power (kW)	286	103	408	84
ω at peak power (rpm)	6000	6300	5900	4400
Peak torque (N m)	531	174	750	209
ω at peak torque (rpm)	4000	4300	4050	2800
Engine volume (L)	5.46	1.80	7.30	2.68
Compression ratio	10.7:1	10.5:1	10:1	9.25:1
Throttle diameter (mm)	50*	60	80*	58

*Value is estimated.

information regarding parameter values used in this study is provided in [Appendix 19.B](#).

The primary criterion to evaluate these models is their accuracy. [Figures 19.1–19.4](#) illustrate the relative performance of the three models with use of the empirical engine data as a benchmark. Each figure pertains to one of the engines and consists of two plots—one for power and the other for torque. In principle, the torque curve should contain the same information as the power curve because power is simply the product of torque and engine speed. However, many empirical torque curves exhibit some differences from what their expected form, so both power and torque curves are included here for complete information.

In [Figure 19.1](#), model II fits the empirical power curve very well. Model III also fits well except for the peak power. Model I meets the peak power but overestimates the remaining part of the empirical curve. In terms of torque, model II meets the peak torque but generally falls under the empirical curve. Model III would give a better fit if it were shifted slightly to the left. Model I generally deviates from the empirical curve by a shift to the left and translation upward.

[Figure 19.2](#) generally shows about the same pattern as that in [Figure 19.1](#), with more noticeable deviations for models I and III. Though model II agrees with the peak torque, the model does not fit the empirical torque curve well under very low and very high engine speeds.

In [Figure 19.3](#), model II generally fits the empirical curves well except for the depressed parts under low to middle engine speeds. Model III's torque curve drops too fast after the peak torque. Model I increasingly deviates from the empirical curves as engine speed decreases.

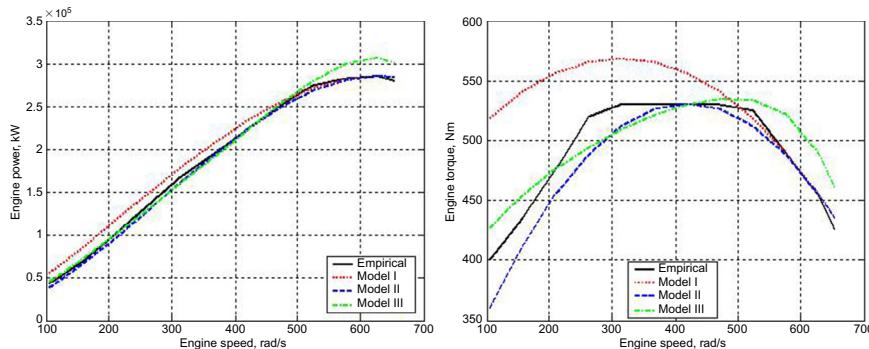


Figure 19.1 Model comparison based on the 2008 Mercedes CLS engine.

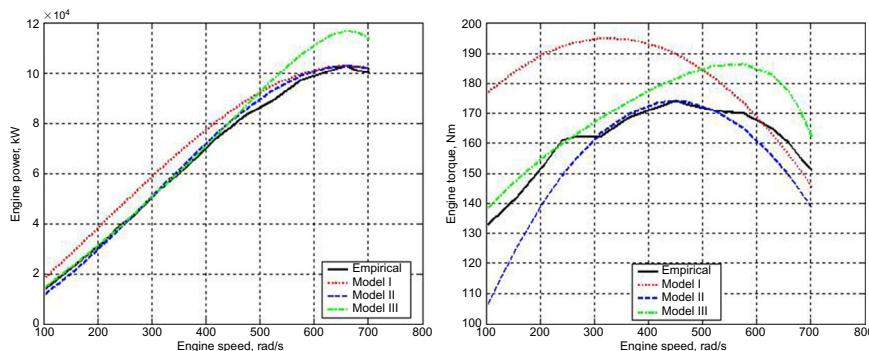


Figure 19.2 Model comparison based on the 2006 Honda Civic engine.

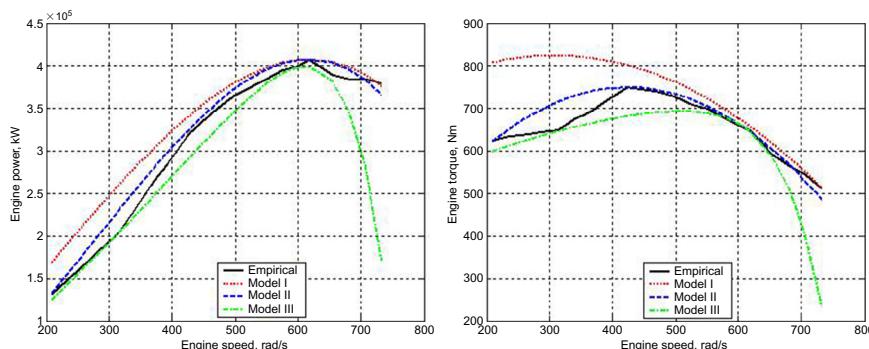


Figure 19.3 Model comparison based on the 2006 Pagani Zonda engine.

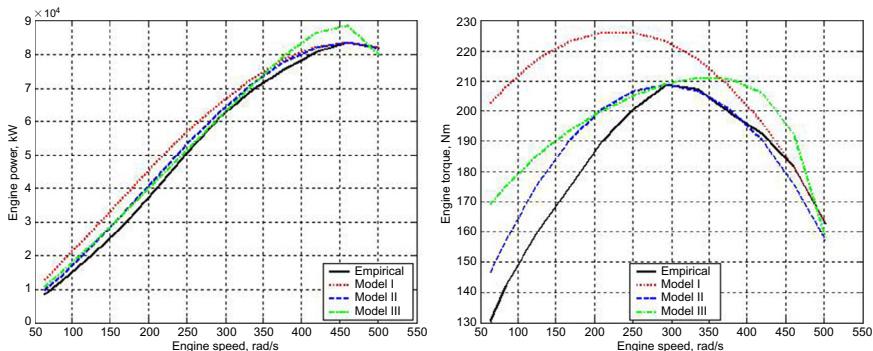


Figure 19.4 Model comparison based on the 1964 Chevrolet Corvair engine.

In Figure 19.4, model II generally overestimates the torque before the peak torque. Except for a good fit of the peak torque, models I and III generally overestimate the torque.

To quantify the accuracy of the three models, the mean absolute percentage error (MAPE) is used as the measure of effectiveness. The MAPE is calculated as

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - X_i|}{Y_i}, \quad (19.27)$$

where n is number of samples, X_i is the model estimate, and Y_i is the corresponding empirical value. Figure 19.5 confirms that model II performs consistently well in both power and torque across the four engines. Its MAPE generally ranges between 3% and 7%. Though less well than model II, model III generally performs quite well, and its MAPE ranges between 4% and 9%. Model I performs the least well of the three models, and its MAPE can be as high as 18%.

The second criterion to evaluate these models is accessibility—that is, the involvement of proprietary parameters and difficult-to-measure variables. In this regard, models I and II are excellent because all they need are peak power and torque and the associated engine speeds. Such information is readily available on the Internet. Model III requires the throttle body diameter, a proprietary parameter, which is less desirable. The third criterion is computational efficiency/model complexity. On average, models I and II consume about 3.2×10^{-5} CPU time to complete a run, while model III takes 0.075 CPU time. Though these numbers appear negligible, the difference is pronounced in real-time applications, especially

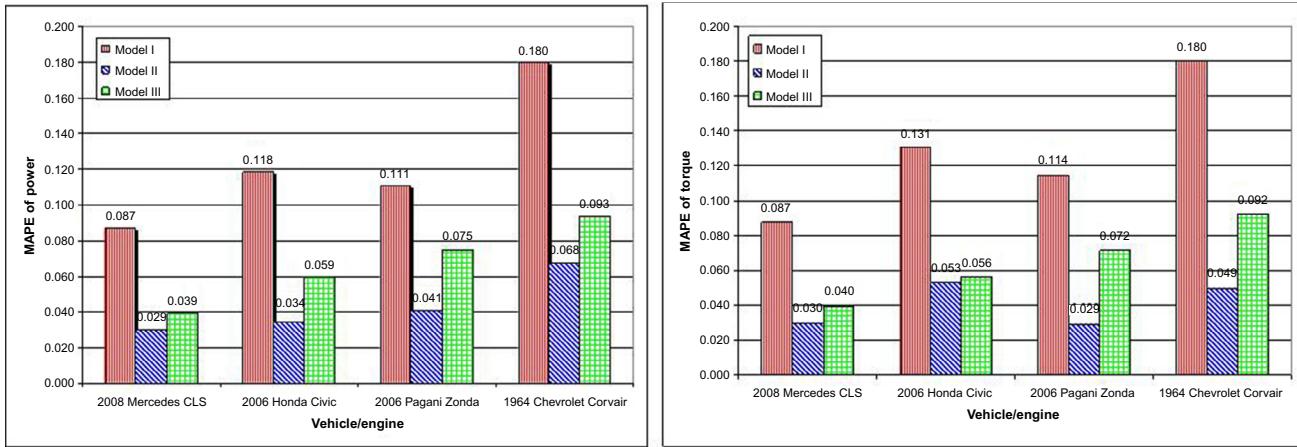


Figure 19.5 MAPE of the three models (left, power; right, torque).

where a procedure has to be repeated very frequently. In terms of the fourth criterion—formulation—all three models are analytical, so no lookup table is involved. The fifth criterion is the need for calibration. In this regard, models I and II involve minimal calibration—all they need are peak power and torque and the associated engine speeds. Calibration of model III is quite involved owing to its proprietary parameter and calibration coefficients. The above comparison results are also highlighted in [Appendix 19.A](#). Overall, model II outperforms the other two models in terms of the above-mentioned evaluation criteria.

19.5 CONCLUSION

An ideal engine model suitable for in-vehicle applications such as a cooperative driving assistance system is expected to have reasonable accuracy, excellent computational efficiency, high accessibility, an analytical formulation, and little need for calibration. Toward these goals, this chapter has presented three simple engine models: model I is an existing one, and models II and III were developed by the author. These models were formulated, validated, and evaluated. In terms of accuracy, models II and III have moderate accuracy, while model I has low accuracy. In terms of computational efficiency, the three models are all acceptable, with models I and II being particularly efficient. In terms of accessibility, models I and II are excellent because they do not require any proprietary parameter or difficult-to-measure variable. All three models are equally good in terms of analytical formulation. Model III requires much effort for calibration, while models I and II involve minimal calibration. Overall, model II appears the best among the three models in terms of all the evaluation criteria.

19.A A CROSS-COMPARISON OF ENGINE MODELS

19.B PARAMETER VALUES

Engine efficiency $\eta = 0.29$

Fuel energy density $E_f = 46900000 \text{ J/kg}$

Stoichiometric air-fuel ratio $\lambda = 0.068$

Air density $\rho = 1.29 \text{ kg/m}^3$

Atmospheric pressure $p = 101325 \text{ Pa}$

$\pi = 3.14159$

Heat capacity ratio of ideal gas $k = 1.407$

Molar mass of air $m' = 28.9$

Model	Accuracy	Complexity	Accessibility	Applications
I [102]	Low	Low	High	Vehicle dynamics
II	Moderate	Low	High	CVT-enabled in-vehicle control
III [101]	Moderate	Moderate	Moderate	Vehicle dynamics
[100]	Low	Low	Moderate	Vehicle dynamics
[99]	Moderate	Moderate	Low	Autonomous cruise control
[73–75]	Moderate	Moderate	Low	Engine control analysis
[78, 79]	High	High	Low	Engine control analysis
[80]	High	High	NA	Powertrain controllers and dynamics
[82]	High	High	Low	Electric throttle control algorithm
[83]	High	High	Low	Air-fuel ratio control and speed control
[90]	High	High	Low	Engine diagnostics and control
[91]	High	High	Low	The design procedure for IC engines
[92]	High	Very high	Low	Upfront design of engines for noise and vibration targets

Universal gas constant $R = 8314.5 \text{ (N m)/(mol K)}$

Coefficients in model III, $\alpha = 0.15$ and $\beta = 10$

PROBLEMS

1. State at least three criteria that you would use to evaluate an engine model.
2. Do an Internet search on manufacturer specifications for the 2016 Volvo XC90 engine and find its peak power and torque as well as the corresponding engine speeds. Use the above information to determine the parameters in [Section 19.3.1](#) (model I) and write the specific functional form of power.
3. Use the result from problem 2 to determine the specific functional form of torque and check whether its peak condition matches the manufacturer specification from an Internet search.
4. Use the results of the Internet search in problem 2 to determine the parameters in [Section 19.3.2](#) (model II) and write the specific functional form of power and torque.

5. Verify if the peak power and torque conditions predicted by the model in problem 4 indeed match those of the manufacturer's specifications.
6. Use the results of the Internet search in problem 2 as well as information in [Appendices 19.A](#) and [19.B](#) to determine the parameters in [Section 19.3.3](#) (model III) for the 2016 Volvo XC90 engine. Assume a throttle diameter of 70 mm, engine displacement $V_e = 0.002 \text{ m}^3$, stagnation pressure $p_0 = 101.325 \text{ kPa}$, and stagnation temperature $T_0 = 293.15 \text{ K}$.

CHAPTER 20

Vehicle Modeling

Picoscopic modeling requires explicit models for vehicles that are separated from driver models. As an attempt in this direction, this chapter is devoted to the modeling of individual vehicle dynamics, using a driver's desired acceleration, deceleration, and steering as inputs to determine vehicle dynamic responses, including longitudinal acceleration, lateral acceleration, and yaw velocity. The vehicle model derived herein is called the dynamic interactive vehicle (DIV) model.¹

20.1 OVERVIEW OF THE DIV MODEL

In automotive engineering, there is a wealth of literature discussing dynamic vehicle models. These models typically come with many degrees of freedom and high modeling fidelity. Typical to these models are their applications in vehicle design, handling, and stability, involving one or a few vehicles. Our interest is a dynamic vehicle model which is well suited for the simulation of a network of vehicles. Such an application involves a large number of interacting vehicles, yet demands a modeling fidelity beyond the microscopic level. On this note, those vehicle models in automotive engineering are overqualified given their complexity and high computation costs. Therefore, a DIV model with high computational efficiency and reasonable modeling fidelity is desirable.

The DIV model will be capable of accepting three inputs from its driver: throttle position, brake pedal position, and steering angle. The model will relate each input to a particular driver's desire and represent the desire on a scale of 0 to 1 for the throttle and brake positions, and on a scale from -1 to 1 for the steering angle. Each of these inputs will then play a role in vehicle dynamics to produce vehicle motion. The following subsections present how the DIV model incorporates essential components of vehicle dynamics in order to faithfully model its motion. These components include the engine, the braking system, and the steering mechanism. Details of

¹This chapter is reproduced from [106].

how the DIV model will account for effects due to rolling resistance, air resistance, and gravity are also presented in the following subsections.

20.2 MODELING LONGITUDINAL MOVEMENT

Forces in the longitudinal direction of the DIV model include the forces due to the engine and the braking system, rolling and aerodynamic resistances, and the force due to gravity. The equation of motion for such a vehicle can be derived by use of Newton's second law of motion:

$$\sum F = m\ddot{x} = F_e - F_b - R_a - R_r - R_g, \quad (20.1)$$

where m is the mass of the vehicle (kg), \ddot{x} is vehicle acceleration (m/s^2), F_e is the tractive force produced by the engine (N), F_b is the force produced by the brake (N), R_a is aerodynamic resistance (N), R_r is rolling resistance (N), and R_g is grade resistance (N).

20.2.1 Modeling Acceleration Performance

The engine plays an important role in vehicle acceleration performance. Here we adopt the engine model recommended in Ref. [72] where engine power P and torque Γ are functions of engine speed ω :

$$\Gamma = \Gamma_{\max} - \frac{P_{\max}}{2\omega_p^2(\omega_p - \omega_t)}(\omega - \omega_t)^2, \quad (20.2)$$

$$P = \frac{P_{\max}}{2\omega_p^2}(3\omega_p - \omega_t)\omega - \frac{P_{\max}}{2\omega_p^2(\omega_p - \omega_t)}(\omega - \omega_t)^2\omega, \quad (20.3)$$

where P_{\max} is the maximum engine power achieved at engine speed ω_p and Γ_{\max} is the maximum engine torque achieved at engine speed ω_t . The four parameters for a specific vehicle are publicly available on the Internet. This model automatically guarantees that $\Gamma = \frac{P}{\omega}$.

Using the torque being delivered to the wheel, we can calculate the engine force F_e produced by the engine to promote vehicle motion with the aid of the appropriate final transmission gear ratio N_{ft} , wheel radius r , and mechanical efficiency of the driveline ζ :

$$F_e = \frac{\Gamma N_{ft} \zeta}{r}. \quad (20.4)$$

20.2.2 Modeling Braking Performance

The brake system is represented by equating the force applied to the brake pedal by the driver to the corresponding deceleration of the vehicle. This means of representing the braking ability of a vehicle is as a result of the work presented in Ref. [107]. The objective of this study was to define the brake characteristics within the space bounded by the relationship between brake pedal force and vehicle deceleration, which will lead to acceptable driver-vehicle performance. In essence, this study determined ergonomic properties for brake pedals that would give drivers the most effective control [108]. Therefore, using the results from this study, the DIV model will be able to account not only for the braking performance of the vehicle but also the manner in which the driver interacts with the brake system.

The results of the aforementioned study include several linear relationships which describe the force being applied to the brake pedal and the rate of deceleration of the vehicle. From these relationships, the DIV model will use the proportionality constant to provide optimal pedal force gain. This proportionality constant, 0.021 g/lb, corresponds to the maximum deceleration rate through minimal pedal force. Using this proportionality constant, we will use the following formulation in the DIV model to represent the brake system of a vehicle and the driver's interaction with that system:

$$F_b = d_b p_f W, \quad (20.5)$$

where F_b is brake force (N), d_b is the driver's desire to brake (0-1), and p_f is the pedal-force gain coefficient.

20.2.3 Modeling Aerodynamic Drag

Aerodynamic drag is another force that retards the motion of a vehicle. This force is dependent on atmospheric conditions, the frontal area of the vehicle, A_f , and the velocity at which the vehicle is traveling relative to the wind, v_r . The equation below further describes aerodynamic drag:

$$R_a = \frac{\rho}{2} C_D A_f v_r^2, \quad (20.6)$$

where ρ is the mass density of air (1.2041 kg/m^3) and C_D is the coefficient of aerodynamic resistance.

20.2.4 Modeling Grade Resistance

The force due to gravity is mainly experienced when the vehicle is on an incline. The force due to gravity that is acting on the vehicle is calculated as

$$F_g = W \sin \theta \approx W \tan \theta = WG, \quad (20.7)$$

where θ is the angle of the incline in radians and G is the grade of the incline with positive sign for upgrade and negative sign for downgrade.

20.3 MODELING LATERAL MOVEMENT

The structure used to represent the movement of the DIV model in the X-Y plane was adapted from Ref. [109], which included the formulation of a kinematics framework and a dynamic framework to model a vehicle's motion in a two-dimensional space. The kinematics framework that was presented in Ref. [109] was chosen for the DIV model for two primary reasons: (1) all the pertinent dynamic properties of the vehicle have already been accounted for by other means in the DIV model, and (2) the ease of use with an accurate X-Y position representation.

At the base of the kinematics framework for the two-dimensional representation of vehicle motion is the treatment of the vehicle as a nonholonomic system, which is a system whose state depends on the path taken in order to achieve it. In addition, nonholonomic constraints are employed under the assumption that there is no slippage at the wheels during a turn. The assumption that there is no slippage at the wheels is predominantly applicable to instances of high-speed cornering, as wheel slippage at low speeds is negligible. The general form of the nonholonomic constraint may be represented as

$$\dot{x} \sin(\phi) - \dot{y} \cos(\phi) = 0, \quad (20.8)$$

where \dot{x} and \dot{y} represent the velocities in the x and y directions of the vehicle coordinate system and ϕ is the vehicle orientation with respect to the global X-Y coordinate system. See Figure 20.1 for an illustration of the coordinate system being used and also for the definition of the variables that will be used in the development of the DIV model.

After a few more iterations of Equation 20.8, the velocity of the center of gravity with respect to the global coordinate system is defined as

$$\dot{X} = \dot{x} \cos(\phi) - \dot{y} \sin(\phi) \quad (20.9)$$

$$\dot{Y} = \dot{x} \sin(\phi) + \dot{y} \cos(\phi) \quad (20.10)$$

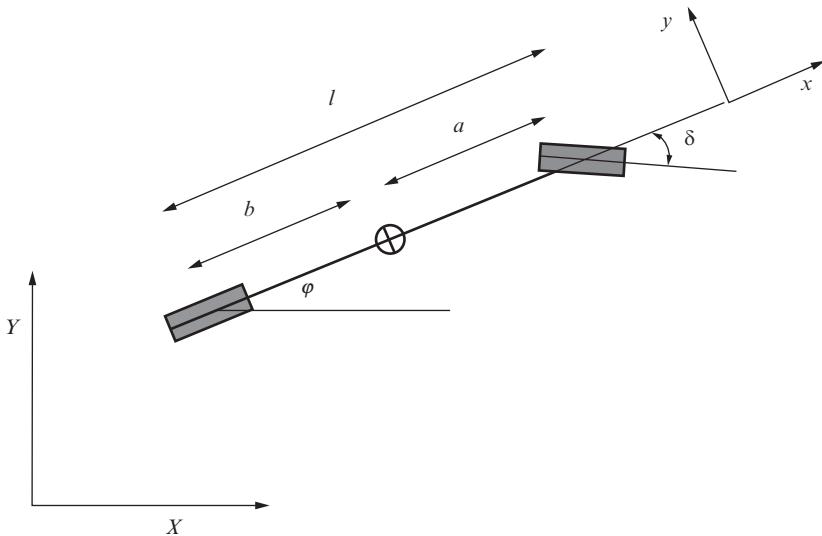


Figure 20.1 DIV model in the X - Y plane.

With Equations 20.9 and 20.10, the global position of the vehicle can now be determined. However, before these equations can be used, lateral velocity, \dot{y} , has to be defined. The definition of the Ackerman angle, δ , also has to be introduced as this is the parameter that is responsible for changing the orientation of the vehicle.

$$\dot{y} = \dot{\phi}b \quad (20.11)$$

$$\dot{\phi} = \frac{\tan(\delta)}{l}\dot{x}, \quad (20.12)$$

and

$$\delta = d_t \frac{\pi N_s}{r_s}, \quad (20.13)$$

where d_t is the driver's desire to turn (-1 to 1), N_s is the number of steering wheel revolutions, and r_s is the steer ratio (ratio of radians turned to the Ackerman angle)

20.4 MODEL CALIBRATION AND VALIDATION

A key feature of the DIV model is that it is meant to be easily calibrated. The calibration of the DIV model will entail the user providing the model with a few performance specifications of the vehicle being modeled. These

specifications will be assessable as they are available to the public via car manufactures and various organizations that offer tools to research a myriad of vehicles—for example, Cars.com. The vehicle performance specifications that the DIV model requires include the aerodynamic resistance coefficient, engine displacement, gear ratios, steer ratio, and the vehicle dimensions.

In addition to these specifications, the model also has a few variables relating to the environment that impact vehicle motion, including wind speed and the gradient of the roadway. Once the values of the vehicle performance specifications and the various values describing the surrounding environment have been entered into the DIV model, it will be able to replicate the motion of the vehicle.

In the validation of the DIV model, three standard performance tests were used to determine whether or not the DIV model is capable of successfully replicating the movement of the vehicle. These tests are typically conducted on vehicles to determine their capabilities of accelerating, braking, and handling. To test vehicle acceleration, the time for a vehicle to go from rest to 97 km/h (60 miles per hour) is recorded, as is the time it takes a vehicle to cover 402 m (a quarter of a mile). The Federal Motor Carrier Safety Administration dictates maximum allowable stopping distances from various speeds that all vehicle manufacturers must satisfy, standardizing vehicle braking. Finally, to measure how well a vehicle handles, the diameter of the circle traced by the vehicle's outer front wheel with the maximum steering angle is recorded.

For details of model calibration and validation, see Ref. [106].

PROBLEMS

1. Conduct an Internet search and find the following information about the 2016 Volvo XC90 engine:
 - a. final drive axle ratio
 - b. first gear ratio
 - c. sixth gear ratio
 - d. tire size
 - e. base curb weight
2. It is known that the vehicle speed v (m/s) is related to the engine speed ω (revolutions per minute) as follows:

$$v = \frac{\pi r}{30N_{\text{ft}}} \omega, \quad (20.14)$$

where r is the tire radius in meters and N_{ft} is the final transmission gear ratio, which is the product of the axle ratio and the gear ratio. Use the information from the previous Internet search and assume the vehicle is cruising at 30 m/s in sixth gear, and find the corresponding engine speed.

3. When a vehicle is starting up, it needs the maximum torque to generate engine force. Assume first gear is used and half of the maximum torque is available at start-up. Calculate the corresponding engine force for a 2016 Volvo XC90 engine assuming the mechanical efficiency of the driveline is 80%.
4. A 2016 Volvo XC90 engine has a drag coefficient C_F of 0.32 and a frontal area A_f of 2.79 m², and is traveling at 100 km/h. How much aerodynamic drag results if the air density ρ is 1.20 kg/m³?
5. The above-mentioned vehicle is running up a hill with a grade G of 5%. Calculate the grade resistance acting on the vehicle.
6. Assume that the above-mentioned vehicle is subject only to aerodynamic drag and grade resistance. Calculate the maximum acceleration available at start-up.

CHAPTER 21

The Field Theory

In picoscopic modeling, drivers are modeled as an intelligent agent who is able to gather information from his or her driving environment and make a decision to achieve his or her goals—for example, traveling to the destination on the preferred route at the desired speed while avoiding hazards. The outputs of the driver model are driving decisions, including steering, accelerating, and braking, which, in turn, can be represented by acceleration in the longitudinal and lateral directions. To serve this purpose, this chapter introduces a generic modeling approach, called the field theory of traffic flow, that represents everything in the environment as a field perceived by the subject driver whose mission is to achieve his or her goals by navigating through the overall field.¹

21.1 MOTIVATION

Research on highway traffic flow over the past half century has resulted in many follow-the-leader theories, each of which was proposed with its own motivation. For examples, in the General Motors family of models [55, 56], a driver's response (e.g., desired acceleration or deceleration) was the result of stimuli (e.g., spacing and relative speed) from his or her leader; the Pipes model [52], the Forbes model [53, 54, 66], and the Gipps model [57] were inspired by safe driving rules; in psychophysical models [64, 111], driver reactions were triggered by perception thresholds; rule-based models [67, 112] were motivated by the fuzzy logic in driver decision making. Though the motivation behind some other car-following models such as the Newell nonlinear model [58] and equilibrium traffic flow models such as those in Refs. [9–12] might not be clear, they were so formulated because of their reasonable performance. Two questions naturally arise. First, would it be possible to have a unifying framework that coherently interprets and relates these models? Second, would it be possible to root such a unifying framework in first principles so that traffic flow theory

¹ This chapter is reproduced from [110].

is furnished with a solid foundation and connected to other branches of sciences and engineering?

This chapter and Chapters 23 and 24 are motivated by the above questions. This chapter attempts to address the second question; Chapter 23 is intended for the first question; Chapter 24 presents a multiscale modeling perspective. In this chapter, our attention shall be devoted to the modeling of driver operational control in a transportation system—that is, the motion and interaction of driver-vehicle units on a long homogeneous highway. From first principles (e.g., physical laws and social rules), a phenomenology is postulated which represents the driving environment perceived by a subject driver as an overall field. In this field, objects (e.g., roadways and vehicles) in the environment are each represented as a component field, and their superposition represents the overall hazard that the subject driver tries to avoid. Hence, the modeling of vehicle motion is simply seeks the least hazardous route by navigating through the overall field along its valley.

21.2 PHYSICAL BASIS OF TRAFFIC FLOW

Three systems are of particular interest: a physical system, a transportation system, and a social system, as illustrated in [Figure 21.1](#). The physical system typically consists of nonliving objects whose motion and interaction are subject to physical laws such as Newton's laws of motion. In contrast, the social system involves living entities such as humans whose behaviors differ widely among the population but generally follow some loosely defined rules (e.g., seeking gains and avoiding losses). As such, physical science is recognized as "hard" since it is more objective, rigorous, and accurate, while social science is perceived as "soft" because of its subjectivity, vagueness, and inexactness. Straddling the above two systems is the transportation system, which involves both living entities (human drivers) and nonliving objects (roadways and vehicles). Hence, transportation science can be perceived as "firm" (for the lack of a proper word between "hard" and "soft") since it deals with both physical laws and social rules. Actually, it is close to the "soft" end when strategic planning is concerned, while it migrates toward the "hard" end if tactical decisions and particularly operational control are of interest.

Many traffic flow phenomena are similar to those in the physical system, yet the transportation system has something special to distinguish itself. Some examples of such similarities are given below.

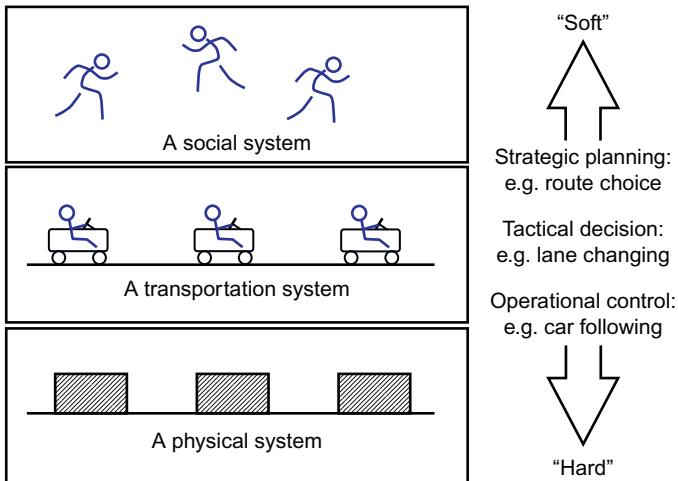


Figure 21.1 Three systems.

21.2.1 Mechanics Phenomena

In physics, forces are the cause of a change of motion. In addition, they are measurable and their effects are reproducible. For example, Newton's second law of motion stipulates that the velocity of an object changes if it is subject to a nonzero external force; Newton's third law says that for every action there is an equal and opposite reaction. Similarly, "forces" exist in traffic flow, but such forces are subjective matters. Consequently, they are nonmeasurable, and their effects do not repeat precisely. For example, a fast driver feels a "force" (a stress in the driver's mind) when he or she approaches a slow vehicle, and hence needs to slow down or change lane. In return, the slow driver may or may not be subject to the reaction "force" depending on whether or not the driver pays attention and responds to the force. If the driver does so, he or she speeds up or gives way in response. Otherwise, Newton's third law does not take effect in this case. More examples of mechanics phenomena are provided below:

M1: Directional flow

Traffic always flows in a predetermined direction much like free objects always fall to the ground. Free objects fall because they are constantly subject to Earth's gravity. Similarly, it is reasonable to imagine that vehicles in the traffic are subject to a "gravity" along the roadway. Such a roadway gravity is, again, a subjective matter since it exists in the mind of drivers and is not

measurable, but it is recognized that the gravity is related to factors such as driver personalities (e.g., aggressiveness), vehicle properties (e.g., engine power), and road conditions (e.g., freeways versus streets).

M2: Free flow

An object in free fall will accelerate to an equilibrium speed because of air resistance, and so does a vehicle in free flow. In this case, the “resistance” comes from the driver’s willingness to comply with traffic rules (e.g., speed limits) as opposed to rolling, grade, and air resistances. Unlike the free-fall speed, which is deterministic and replicable given the same condition, the free speed of a vehicle is, once again, a subjective matter because it is largely the driver’s choice. Given the same conditions, the choice may differ for different drivers and, for the same driver, at different times. In addition, different roadways support different free speeds. To avoid confusion, the free speed chosen by a driver is termed his or her “desired speed,” whereas the free speed aggregated over a group of vehicles on a particular road is called the “free-flow speed” supported by the road. Generally, the desired speed is related to driver personalities and road conditions, while the free-flow speed is affected by road conditions and the driver population.

M3: Stopping at a red light

Much like a moving object being slowed to a stop behind a wall, a vehicle decelerates to a stop in front of a red light. The analogous “repelling force” in the latter case resides in the driver in that if he or she ignores the red light, the consequence is costly (e.g., an accident or a ticket). Unlike the moving object, which always stops in the same fashion in repeated experiments, drivers are entitled to decelerate at a comfortable rate to a stop and, in some extreme cases, drivers may forget to stop.

M4: Road barriers

Vehicles moving in the same direction on a roadway are separated by lane lines. To avoid colliding with vehicles in adjacent lanes, a driver must keep in his or her lane as if he or she were guided by barriers at both edges of the lane. If, however, the driver unconsciously departs from the current lane, he or she will perceive some stress, which motivates him or her to steer back into the lane as if a correction “force” from the barrier acts on the vehicle and pushes it toward the center of the lane. If the driver is blocked by a slow vehicle, the desire for mobility will motivate the driver to change lane as if he or she were energized or elevated above the barrier so he or she can cross it and land on the adjacent lane. Running off the road is discouraged,

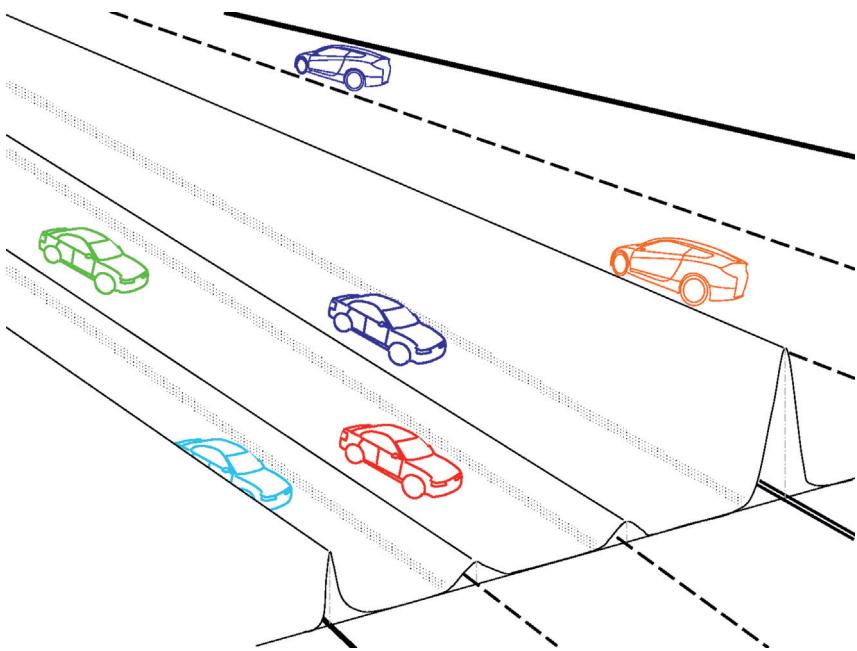


Figure 21.2 Road barriers.

so barriers at road edges are typically higher than lane barriers. Encroaching into the opposite direction of travel is so dangerous that the barrier at the center line is very high (see [Figure 21.2](#) for an illustration of the barriers). These barriers are not real objects, but are only imaginary in drivers' minds.

21.2.2 Electromagnetic Phenomena

An object can exert a force on another object in either of the following ways: collision and action at a distance. For example, hitting a ball with a bat is an example of the former and finding a needle with use of a magnet is an example of the latter. Though collisions are not uncommon on highways, action at a distance is how vehicles normally interact with each other, and examples of this kind include some of the above-mentioned mechanics phenomena as well as the following:

E1: Car following

When a fast vehicle catches up with a slow vehicle, the fast driver perceives an imminent collision if he or she keeps driving at the high speed. The cost and fear of the collision motivates the fast driver to take actions in advance.

If a lane change is not an option and the slow driver does not speed up, the fast driver has to decelerate when he or she approaches the slow vehicle, and then adopts the slow vehicle's speed separated by a safe following distance. This is analogous to moving a charge A toward a like charge B. According to Coulomb's law, the electric force between them is directly proportional to the product of their charges and inversely proportional to the square of their distance. Similarly in car following, the "force" (stress) acting on the fast driver is larger if he or she runs into the slow vehicle faster and their separation is shorter. However, the same opposite force may or may not act on the slow driver as he or she may or may not notice the vehicle approaching from behind.

E2: Tailgating

We continue with the above example and assume that the fast vehicle tailgates (i.e., follows at a dangerously short distance). Then, it is likely that the opposite force is perceived by the slow driver, who may respond by speeding up or giving way to the fast follower. We return to the analogy, but charge B is now driven (or driven away) by charge A and Newton's third law holds in this case. In general, a "force" must be perceived by a driver before the force has an effect on the person. In addition, a driver's ability to perceive something depends on where he or she scans and how frequently this happens.

E3: Shying away

If two vehicles happen to run in parallel, one or both drivers may feel intimidated. The fear of a side collision motivates them to spread out in space (longitudinally or laterally). Such a shying-away effect becomes more evident when one of the vehicles is a heavy truck.

21.2.3 Wave Phenomena

W1: Harmonic wave

A platoon of vehicles on a roadway is like a harmonic wave. The platoon is characterized by flow (in vehicles per hour), traffic speed (in kilometers per hour), and density (in vehicles per kilometer), while the wave is determined by frequency (in hertz, or cycles per second), wave speed (in meters per second), and wave length (in meters). One immediately recognizes that flow is equivalent to frequency, traffic speed is equivalent to wave speed, and the spacing (the inverse of density) is equivalent to wave length. The upper part of [Figure 21.3](#) shows a platoon of vehicles as a harmonic wave.

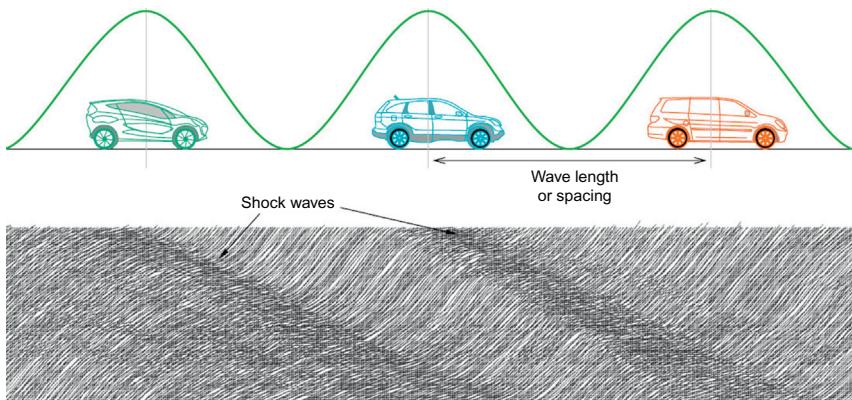


Figure 21.3 Traffic and waves.

W2: Signal propagation

The signal here does not mean a traffic signal, rather it refers to any quantity that clearly defines the location and speed of a perturbation in a medium. When the leading vehicle of a compact platoon brakes briefly, a kinematic wave forms and propagates against the platoon, where the signal here is the brief speed reduction. When a platoon of fast vehicles catches up with a platoon of slow vehicles, a shock wave is generated and propagates against the traffic, where the signal here is the interface between fast and slow vehicles. The bottom part of Figure 21.3 illustrates a few shock waves observed in vehicle trajectories.

W3: Wave-particle duality

All matter, particularly small-scale objects, exhibits both wavelike and particle-like properties. The latter is prominent when individual objects are concerned (e.g., the photoelectric effect), while the former becomes significant when the behavior of many objects is viewed collectively (e.g., diffraction of waves). In traffic flow, individual vehicles act like particles (e.g., car following and lane changing), while a platoon or platoons of vehicles act like waves (e.g., kinematic waves and shock waves).

21.2.4 Statistical Mechanics Phenomena

Traffic flow has been modeled by many authors as a one-dimensional compressible fluid, such as a gas. In gases, the speeds of gas molecules follow a Maxwell-Boltzmann distribution (see Figure 21.4), top left, for an illustration. Remarkable in the distribution is the increase in average

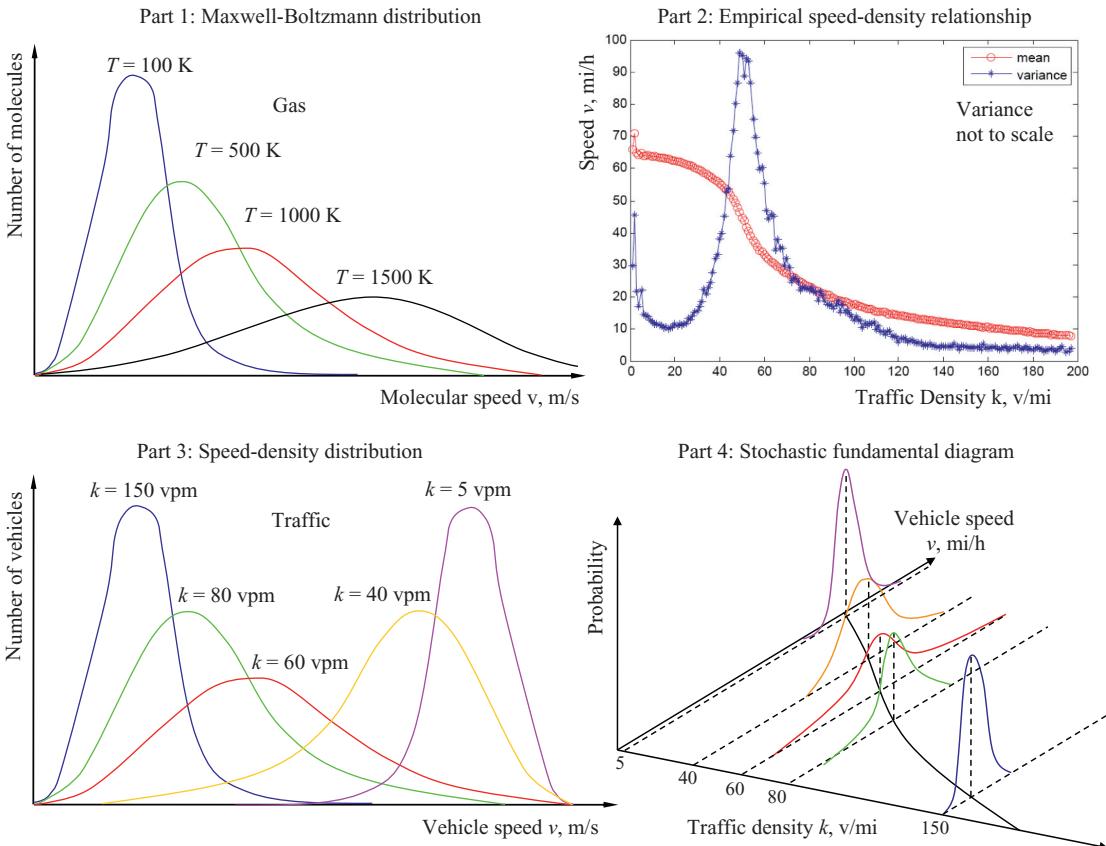


Figure 21.4 Maxwell-Boltzmann distribution in traffic flow.

speed and speed variance as temperature increases. In contrast, traffic flow exhibits a different trend. Empirical observations ([Figure 21.4](#), top right) show that the variance of traffic speed peaks around the optimal density (where capacity flow occurs) and drops at both ends. Such a distribution is illustrated [Figure 21.4](#), bottom left, in contrast to the Maxwell-Boltzmann distribution, and is further elaborated in a three-dimensional model in [Figure 21.4](#), bottom right, which forms the basis of a stochastic fundamental diagram.

21.3 THE FIELD THEORY

The above-mentioned similarities between the transportation system and the physical system provide motivation for a phenomenology of traffic flow (i.e., the field theory) which aims to describe traffic phenomena in a way that is consistent with first principles but is not directly derived from them. Since the transportation system involves both living entities (e.g., human drivers) and nonliving objects (e.g., roadways and vehicles), it is subject to both physical laws and social rules. As such, the phenomenology is formulated progressively on the basis of a set of postulates, two of which (Postulates 1 and 3) are physical and two of which (Postulates 2 and 4) are social.

21.3.1 Postulate 1: A Road is a Physical Field

Postulate 1 is motivated by phenomena M1, M2, and M4 in [Section 21.2](#). In the longitudinal (x) direction, a driver-vehicle unit is subject to a gravity along the road:

$$G_i = m_i \times g_i, \quad (21.1)$$

where i denotes the unit's ID, G_i is the roadway gravity acting on the unit, m_i is the mass of the unit, and g_i is the acceleration of roadway gravity perceived by driver i . As discussed in M1, g_i is a function of driver personalities Θ , vehicle properties Λ , and road conditions Ξ —that is, $g_i = g_i(\Theta, \Lambda, \Xi)$.

Meanwhile, the unit is also subject to a resistance R_i perceived by the driver due to his or her willingness to observe traffic rules (e.g., speed limits). As discussed in M2, R_i is related to the driver's perceived difference between his or her actual speed \dot{x}_i and the desired speed v_i , which in turn is related to the free-flow speed of the road v_f —that is, $R_i = R_i(\dot{x}_i, v_i, v_f)$. Therefore, the net force acting on unit i in the longitudinal direction can be expressed as

$$m_i \ddot{x}_i = G_i - R_i, \quad (21.2)$$

where \ddot{x}_i is the acceleration of unit i . Since the right-hand side represents the amount of net force that can be used to accelerate the unit, it can be interpreted as the driver's *unsatisfied desire for mobility*. As the unit speeds up, the right-hand side decreases (because R_i increases). Eventually, the right-hand side vanishes, at which time the unit reaches its desired speed v_i . If, somehow, a random disturbance brings the unit's speed above v_i , the right-hand side becomes negative. In this case, the unit decelerates and finally settles back to v_i .

In the lateral direction of the road, there are lane lines, road edges, and center lines to guide and separate traffic. As discussed in M4 in [Section 21.2](#), these cross-section elements of the road can be mapped into a roadway potential field U_i^R perceived by the driver. When the unit deviates from its lane, the unit is subject to a correction force N_i , which can be interpreted as the stress on the driver to keep in his or her lane (see [Figures 21.2](#) and [21.6](#)). The effect of such a force is to push the unit back to the center of the current lane. On the basis of physical principles, the force can be determined as the derivative of the roadway field, U_i^R , with respect to the unit's lateral displacement y_i :

$$N_i = -\frac{\partial U_i^R}{\partial y_i}. \quad (21.3)$$

21.3.2 Postulate 2: A Driver Responds to His or Her Surroundings Anisotropically

The interaction between two driver-vehicle units differs from the collision of two objects in two ways: one pertains to Newton's third law of motion, which is discussed below, and the other concerns noncontact forces, which are the subject of the next postulate.

In classical physics, Newton's third law of motion holds when two objects collide with each other. However, the law generally does not hold in the interaction between two driver-vehicle units. For example, when a fast vehicle catches up with a slow vehicle, the fast driver perceives a “repelling force” (i.e., stress) as the gap closes. The smaller the gap, the greater the force. Conversely, the reaction force may or may not be perceived by the slow driver depending on whether he or she notices the approaching fast vehicle and his or her willingness to respond. Since drivers all sit facing the front, it is the driver behind who is responsible for watching for safe distances and who is held liable for a rear-end collision should it happen. Therefore,

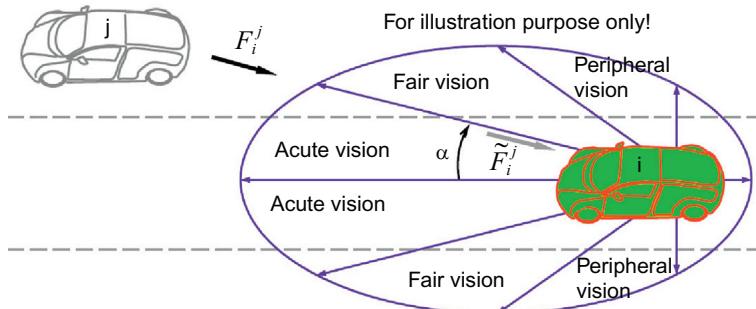


Figure 21.5 Distribution of driver's attention.

it is not uncommon that a leading driver does not respond to situations happening behind such as an approaching fast vehicle.

In general, a driver's responsiveness to his or her surroundings varies with his or her viewing angle and scanning frequency (see Figure 21.5). For example, the area immediately in front of the driver, especially in the same lane, falls into the driver's acute vision zone. The driver is responsible for watching this area constantly and responding to a situation promptly. Roughly in the driver's fair vision zones, the frontal areas in side lanes receive a considerable amount of the driver's attention since vehicles in the side lanes may change to the subject lane and the driver needs to watch this area when changing lanes. In comparison, the driver scans less frequently at both sides of his or her vehicle (roughly the driver's peripheral vision zones) unless the driver needs to change lanes or avoid parallel running. The last and least attended area is the rear of the vehicle, not only because it is difficult to access (indirectly by means of side or rear mirrors), but also because liability rests with drivers behind. Therefore, it is reasonable to assume that the driver's directional response to his or her surroundings, γ_i , is a function of his or her viewing angle α_i —that is, $\gamma_i = \gamma(\alpha_i)$. Consequently, the force that actually acts on the unit, \tilde{F}_i , is the product of the force that might have been perceived by the driver if he or she had paid full attention to it, F_i , and his or her directional response γ_i —that is,

$$\tilde{F}_i = F_i \times \gamma(\alpha_i), \quad (21.4)$$

where $\alpha_i \in [-\pi, \pi]$ is the viewing angle. For example, if one chooses $\gamma(0) = 1$ and $\gamma(\pi) = 0$, the driver responds to F_i in full when it comes from a leading vehicle (i.e., $\alpha_i = 0$) and ignores F_i when it comes from a trailing vehicle (i.e., $\alpha_i = \pi$).

21.3.3 Postulate 3: A Driver Interacts with Others by Action at a Distance

As described in E1, E2, and E3 in [Section 21.2](#), a driver is able to sense the presence of other vehicles and obstacles in his or her vicinity and take preventive actions to avoid a collision. It is postulated that such an action at a distance is mediated by a field which is perceived by the driver as the danger of a collision. One may imagine the field as a hill; the higher and steeper the hill is, the more difficult it is to climb. The base of the hill/field delineates a region, outside of which the driver is not influenced by the field. For example, the dash-dotted oval (labeled as “Base j ”) in [Figure 21.6](#) represents the base of the field perceived by driver i due to unit j . One may also interpret the field as the personal space of unit j , into which intrusion is discouraged. The deeper unit i intrudes, the stronger the repelling force it receives. The longitudinal section of the field is illustrated as the curve above the x -axis.

Similarly, unit k represents another field (whose base is labeled as “Base k ”) which also exerts an influence on unit i . Since unit k is in the lane at

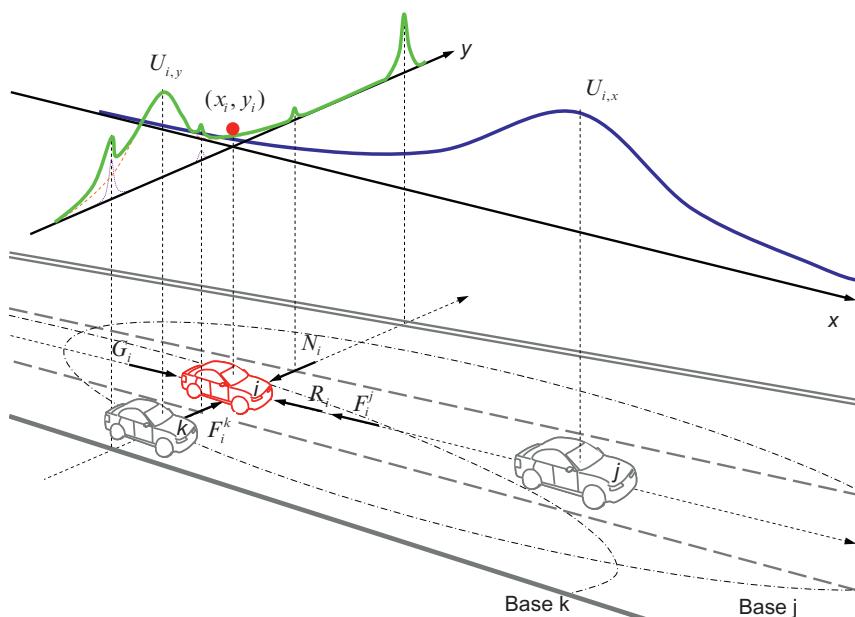


Figure 21.6 The illustration of a field.

the side of unit i , the influence is not in the longitudinal x direction but is in the lateral y direction—that is, the field results in a repelling force, F_i^k , on unit i , which motivates it to shy away from unit k .

The above fields and, consequently, forces are related to driver personalities Θ and vehicle dynamics Γ —that is, $U = U(\Theta, \Gamma)$ and $F = F(\Theta, \Gamma)$. For example, since an aggressive driver accepts shorter car-following distances, the field perceived by such a driver covers a smaller base. On the other hand, the faster a unit moves, the more hazard it imposes on neighboring vehicles, and thus the larger and steeper is the field it creates.

21.3.4 Postulate 4: A Driver Tries to Achieve Gains and Avoid Losses

A driver's strategy of moving on roadways is to achieve mobility and safety (gains) while avoiding collisions and violation of traffic rules (losses). Such a strategy can be represented with use of an overall potential field U_i which consists of component fields such as those due to moving units U_i^B , roadways U_i^R , and traffic control devices U_i^C —that is,

$$U_i = U_i^B + U_i^R + U_i^C. \quad (21.5)$$

If U_i is viewed as a mountain range whose elevation denotes the risk of losses, the driver's strategy is to navigate through the mountain range along its valley—that is, the least stressful route. For example, Figure 21.6 illustrates two sections of such a field. Perceived by driver i , the longitudinal x section of the field, $U_{i,x}$, is dominated by unit j since it is the only neighboring vehicle in the center lane. Unit i is represented as a ball which rides on the tail of curve $U_{i,x}$ since the vehicle is within unit j 's field. Therefore, unit i is subject to a repelling force F_i^j which is derived from $U_{i,x}$ as

$$F_i^j = -\frac{\partial U_{i,x}}{\partial x}. \quad (21.6)$$

The effect of F_i^j is to push unit i back to keep a safe distance. By incorporating the driver's unsatisfied desire for mobility ($G_i - R_i$), we can determine the net force in the x direction as:

$$m_i \ddot{x}_i = \sum F_{i,x} = G_i - R_i - F_i^j = (m_i g_i - R_i) + \frac{\partial U_{i,x}}{\partial x}. \quad (21.7)$$

The section of U_i in the lateral y direction, $U_{i,y}$ (the bold curve), is the sum of two components: the cross section of the field due to unit k (the dashed curve) and that due to the roadway field (the dotted curve).

The former results in a repelling force F_i^k which makes unit i to shy away from unit k and the latter generates a correction force N_i should unit i depart its lane center. Therefore, the net effect can be expressed as:

$$m_i \ddot{y}_i = \sum F_{i,y} = F_i^k - N_i = -\frac{\partial U_{i,y}}{\partial y} \quad (21.8)$$

By incorporating time t , driver i 's perception-reaction time τ_i , and driver i 's directional response y_i , we can express Equations 21.7 and 21.8 as

$$m_i \ddot{x}_i(t + \tau_i) = \sum \tilde{F}_{i,x}(t) = \gamma_i^0 [G_i(t) - R_i(t)] + \gamma(\alpha_i^j) \frac{\partial U_{i,x}}{\partial x}, \quad (21.9a)$$

$$m_i \ddot{y}_i(t + \tau_i) = \sum \tilde{F}_{i,y}(t) = -\gamma(\alpha_i^k) \frac{\partial U_{i,y}}{\partial y}, \quad (21.9b)$$

where $\gamma_i^0 \in [0, 1]$ represents the unit's attention to its unsatisfied desire for mobility (typically $\gamma_i^0 = 1$), and α_i^j and α_i^k are viewing angles, which are also functions of time. The above system of equations summarizes the field theory in generic terms and constitutes the basic law governing a unit's motion on a planar surface.

21.4 SIMPLIFICATION OF THE FIELD THEORY

Though the generic form of the field theory is able to explain some traffic phenomena qualitatively, rigorous modeling of traffic flow requires a specific form, which is the focus of this section and the next chapter. In the generic theory, the functional forms of the field U_i , roadway gravity G_i , and resistance R_i are undetermined. It appears that the generic theory can take many specific forms, and it is impractical to enumerate all of them. In choosing a specific form, we find that Occam's razor turns out to be a good rule of thumb, and basically says that "entities should not be multiplied unnecessarily." Hence, the razor gives rise to the following considerations: (1) the chosen specific form should make physical sense, for which empirical observations are good tests, (2) it should take a simple functional form that involves physically meaningful parameters but not calibration coefficients, and (3) it should provide a sound microscopic basis for aggregated behavior—that is macroscopic equilibrium models. With these considerations in mind, some simplifications are made to the generic theory as the first step in the formulation of a specific form.

Simplification 1

Rather than formulating the field itself, the specific form formulates forces resulting from the field directly.

Simplification 2

The specific form decouples equations in the longitudinal x direction and lateral y direction—that is, the longitudinal equation is used to model driver's longitudinal control (e.g., car-following behavior) and the lateral equation is used only when a lane change is to be considered.

Simplification 3

Directional response $\gamma(\alpha)$ is treated as follows: for car following, the subject driver responds only to his or her leader; for lane changing, the subject driver responds to the leading and trailing vehicles in the current lane and the target lane.

21.4.1 Motion in a Longitudinal Direction

With these simplifications, vehicle motion in the longitudinal direction is formulated as follows. Note that time t and response delay (i.e., perception-reaction time τ) are dropped for convenience.

The term $(G_i - R_i)$ explains a driver-vehicle unit's unsatisfied desire for mobility. Intuitively, when a unit starts from standstill—that is, $\dot{x}_i = 0$ —its unsatisfied desire for mobility is the greatest. As the unit speeds up, $(G_i - R_i)$ decreases accordingly, but is still positive—that is, it still accelerates the unit to higher speeds. When the unit achieves its desired speed—that is, $\dot{x}_i = v_i$ —its desire for mobility has been fully satisfied and, hence, $G_i - R_i = 0$, which means that the unit settles at v_i if no other forces act on it. If a random perturbation brings \dot{x}_i over v_i , the unit's desire for mobility is oversatisfied and $G_i - R_i$ becomes negative, which decelerates the unit back to v_i . With the above understanding, a specific form of the unsatisfied desire for mobility can be formulated as

$$G_i - R_i = m_i g_i \left[1 - \left(\frac{\dot{x}_i}{v_i} \right)^\delta \right], \quad (21.10)$$

where δ is a calibration parameter.

When a fast unit i (with displacement x_i , speed \dot{x}_i , and acceleration \ddot{x}_i) catches up with a slow unit j (with x_j , \dot{x}_j , and \ddot{x}_j), the former is subject to a noncontact force, F_i^j , from the latter. Such a noncontact force varies with the spacing between the two units, $s_{ij} = x_j - x_i$. For example, the force

virtually has no effect on unit i when it is distant, but has an effect when unit i becomes close (e.g., within the range of its desired spacing s_{ij}^*), increases as the spacing becomes even shorter ($s_{ij} \downarrow$), and it goes to a maximum when $s_{ij} \rightarrow l_j$, where l_j represents the minimum “safety room” required by unit j , an extreme case of which is the length of vehicle j . In addition, the effect of the force is also related to the speeds and relative speed of units i and j . Such an effect can be incorporated into the formulation of driver i ’s desired spacing s_{ij}^* .

Therefore, a simple way to represent the force is to use an exponential function. The general idea of this model is to set the desired spacing s_{ij}^* as a baseline, beyond which the intrusion by unit i is translated exponentially to the repelling force acting on the unit. As such, a more specific but still quite generic form of the force is

$$F_i^j = f(e^{s_{ij}^* - s_{ij}}), \quad (21.11)$$

where $s_{ij}^* - s_{ij}$ represents how far unit i intrudes into s_{ij}^* .

Combining the above, we can express the effort that is required by driver i to control his or her vehicle in the longitudinal direction as

$$m_i \ddot{x}_i = G_i - R_i - F_i^j = m_i g_i \left[1 - \left(\frac{\dot{x}_i}{v_i} \right)^\delta \right] - f(e^{s_{ij}^* - s_{ij}}) \quad (21.12)$$

or

$$m_i \ddot{x}_i = m_i g_i \left[1 - \left(\frac{\dot{x}_i}{v_i} \right)^\delta - f(e^{s_{ij}^* - s_{ij}}) \right] \quad (21.13)$$

if the coefficient of F_i^j is chosen properly. Though Equation 21.13 can be instantiated in many possible ways, the following special case is of particular interest. Putting time t and response delay τ back in and eliminating vehicle mass m from both sides, we find Equation 21.13 can take the following special form:

$$\ddot{x}_i(t + \tau_i) = g_i \left[1 - \left(\frac{\dot{x}_i(t)}{v_i} \right)^\delta - e^{\frac{s_{ij}(t)^* - s_{ij}(t)}{Z}} \right]. \quad (21.14)$$

If one chooses $\delta = 1$ and $Z = s_{ij}(t)^*$, the above equation reduces to a more specific form:

$$\ddot{x}_i(t + \tau_i) = g_i \left[1 - \left(\frac{\dot{x}_i(t)}{v_i} \right) - e^{\frac{s_{ij}(t)^* - s_{ij}(t)}{s_{ij}(t)^*}} \right]. \quad (21.15)$$

This model will be further elaborated and analyzed in the next chapter.

21.4.2 Motion in a Lateral Direction

A driver-vehicle unit's motion in the lateral γ direction involves a decision at two levels: lane change and gap acceptance. A lane-change decision concerns the driver's desire to change to an adjacent lane to better achieve his or her goals such as mobility and safety, a situation which typically happens when the driver is blocked by a slow leader in the current lane. A gap-acceptance decision addresses the execution of the lane change decision by physically moving the vehicle into the target lane when an opportunity comes up (e.g., a safe gap is available in the target lane). Figure 21.8 illustrates the scenario in Figure 21.7 from a different and broader perspective involving the subject unit i and its leader j in the right lane and a trailing neighbor p in the left lane. The top part of Figure 21.8 illustrates unit

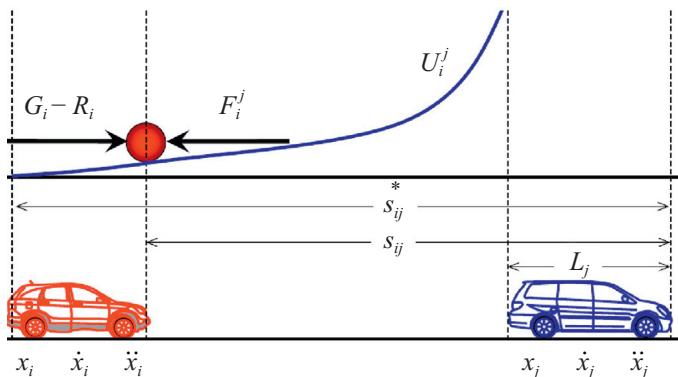


Figure 21.7 Action at a distance.

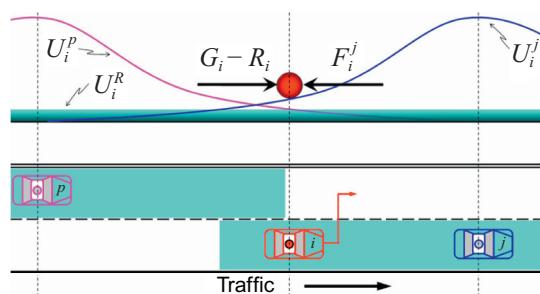


Figure 21.8 Lane change.

i (the ball) in its perceived potential fields: U_i^j due to unit j in the same lane, U_i^p due to unit p in the left lane, and U_i^R due to the road barrier (lane line). Driven by its desire for mobility, unit i climbs up onto U_i^j , during which time unit i has to adapt to unit j 's speed while achieving a balance between $(G_i - R_i)$ and F_i^j . Under this circumstance, unit i reaches its decision on a lane change in order to satisfy its desire for mobility. With this decision, driver i begins to seek opportunities in adjacent lanes. In this particular example, the right side is obviously not an option since it is prohibitive to move off the road. Hopefully, an opportunity exists in the left lane because the elevation of unit i (where the ball rides) is higher than both the lane barrier U_i^R and the front of field U_i^p . Therefore, unit i initiates a smooth transition by laterally rolling off the tail of U_i^j , crossing over U_i^R , and landing on the front of U_i^p , the effect of which is shown in the middle part of [Figure 21.8](#).

We can further simplify the above lateral control model by reducing a smooth field to a flat “personal space” into which intrusion by another unit is undesirable. For example, the bottom part of [Figure 21.8](#) illustrates units j and p 's personal spaces after elimination of the lane barrier. A lane-change decision is reached whenever a unit intrudes into another unit's personal space, which certainly applies to unit i . With such a decision, unit i begins to search for open spaces in adjacent lanes, and one happens to be available in the left lane. Hence, the result of the gap-acceptance decision is to abruptly switch unit i to the left lane.

21.5 DISCUSSION OF THE FIELD THEORY

Responding to the two questions posed at the beginning of this chapter, we state the field theory can serve as a unifying framework that is able to coherently relate existing models to each other. This will be the topic of Chapter 23. Meanwhile, the field theory is proposed with its roots in both physical science and social science and, therefore, establishes the foundation that allows transportation to be treated as a science. Such a role of the field theory is discussed further below.

21.5.1 Tentative Definition of Two Vague Terms

On the basis of the field theory, it is possible to quantify two vague terms—namely, mobility and congestion—which are frequently used in the transportation profession without rigorous definition.

Mobility

The dictionary definition of *mobility* is “the quality of moving freely.” As such, the quality reaches 100% if an individual is able to move as he or she desires, while the quality drops to 0 if the person is stuck in a traffic jam. Therefore, the phenomenological interpretation of personal mobility $M_i(t)$ at an instant of time t perceived by driver i can be expressed as the portion of his or her desired speed which has been satisfied—that is,

$$M_i(t) = \frac{\dot{x}_i(t)}{v_i}, \quad (21.16)$$

where $\dot{x}_i(t)$ is driver i ’s actual speed and v_i is the desired speed. Since v_i is (typically) greater than $\dot{x}_i(t)$, personal instant mobility $M_i(t)$ ranges between 0 and 1. The mobility perceived by the same driver over the course of a journey can be represented as the average of $M_i(t)$ over trip time T_i :

$$M_i = \frac{1}{T_i} \int_0^{T_i} \frac{\dot{x}_i(t)}{v_i} dt. \quad (21.17)$$

Again, personal mobility M_i falls between 0 and 1. Similarly, the mobility perceived by all drivers in a traffic system can be calculated as

$$M = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{T} \int_0^{T_i} \frac{\dot{x}_i(t)}{v_i} dt \right), \quad (21.18)$$

where N is number of drivers in the traffic system. Note that system mobility $M \in [0, 1]$ can be used as an indication of the level of service provided by the traffic system and perceived by drivers.

Congestion

The dictionary definition of *congestion* is “a state that is so crowded as to hinder or prevent freedom of movement.” One candidate quantification of congestion, C , can be the opposite of mobility—that is, $C = 1 - M$, which is expressed relatively as a percentage. Another, perhaps more meaningful, way to quantify congestion is to recognize the “stress” experienced by a driver when he or she moves in a traffic system. Therefore, the phenomenological interpretation of personal congestion $C_i(t)$ at an instant of time t can be expressed in absolute terms as the stress (or equivalently hazard or potential) $U_i(t)$ perceived by driver i —that is,

$$C_i(t) = U_i(t). \quad (21.19)$$

As defined in Equation 21.5, the overall potential consists of potentials due to moving units U_i^B , roadways U_i^R , and traffic control devices U_i^C . Therefore, a driver would experience no congestion if he or she moves on a roadway that is free of impedance from other moving units, the need for lane changes, and traffic control devices. Therefore, any increase of these would add to the driver's perception of congestion. Consequently, personal congestion perceived by the same driver over the course of a journey can be represented as the sum of $C_i(t)$ over trip time T_i :

$$C_i = \int_0^{T_i} C_i(t) dt. \quad (21.20)$$

Further, the congestion experienced by all drivers in a traffic system can be calculated as the sum of personal congestion over the driver population:

$$C = \sum_{i=1}^N C_i. \quad (21.21)$$

21.5.2 Connection to the Existing Knowledge Base

The purpose of this discussion is to place the field theory in a broader context and show how the field theory relates to the existing knowledge base and, in return, how successful experience of related fields can be transferred to solve our problems at hand.

Connection to Other Traffic Flow Theories

Existing microscopic traffic flow models emphasize the application of social rules or human factors in the modeling of car-following behavior, whereas the work presented in this chapter attempts to integrate both social rules and physical principles in the modeling of traffic flow. Readers are referred to Chapter 23, where such a connection is elaborated and presented from a unified perspective.

Connection to Other Engineering Disciplines

The Lennard-Jones potential plays an important role in engineering, particularly in granular flow and molecular dynamics. In molecular dynamics, computer simulation is employed to trace the time evolution of a set of interacting particles (e.g., atoms or molecules) by integrating their equations of motion. The Lennard-Jones potential is the underlying model

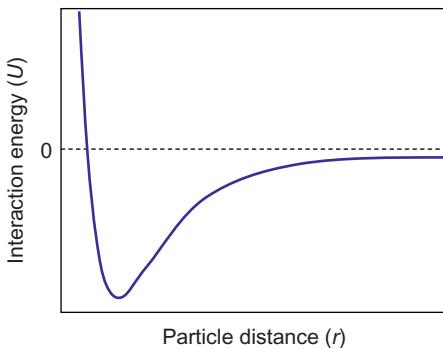


Figure 21.9 Lennard-Jones potential.

to determine the motion and interaction of these particles. In materials engineering and granular flow, the Lennard-Jones potential is typically used as the constitutive law to determine the interaction of two particles. With a clear understanding of the constitutive law of two particles, systems consisting of a large quantity of these particles (e.g., many-body systems) can be simulated and analyzed. Illustrated in Figure 21.9, the Lennard-Jones potential takes the following form:

$$U(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right], \quad (21.22)$$

where U is the Lennard-Jones potential due to particle interaction, r is the distance between two particles, ϵ is the depth of the potential well, and σ is the distance at which the interparticle potential is zero. The equation is actually a superposition of two terms: a long-range attraction term $(\frac{\sigma}{r})^6$ and a short-range repulsion term $(\frac{\sigma}{r})^{12}$.

The phenomenology, in particular Equation 21.12, takes a similar form. For example, the long-range attraction ($m_i \ddot{x}_i = m_i g_i [1 - (\dot{x}_i / v_i)]$) is due to a driver's desire for mobility and the short-range repulsion $f(e^{s_{ij}^* - s_{ij}})$ is due to safety rules. In addition, the interaction between two vehicles is a function of the spacing s_{ij} (equivalent to r) between them, and there is an equilibrium spacing s_{ij}^* (equivalent to σ) around which the attraction equals the repulsion. Therefore, the Lennard-Jones potential in a transportation system can be derived from the phenomenology. With such a bridge, transportation and related engineering disciplines are able to not only learn from but also shed light on each other.

Connection to Physical Science

The phenomenology proposed herein represents a body of knowledge that originates from empirical observations and, in return, that is able to explain real-world phenomena. Rather than being derived directly from first principles, the phenomenology is formulated in a way that is consistent with fundamental theory. For example, the essence of the field theory in the phenomenology is the explanation of an individual driver's action by recourse to his or her position in relation to others. The driver's position in the field in turn gives rise to a force acting on the person, but such a force is motivated from within as opposed to being applied from without. As another example, Equation 21.9 is a special form of Newton's second law of motion if one ignores the driver's perception-reaction time τ and directional response γ . In addition, action at a distance as a means of interaction between drivers becomes a hard collision if the driver's need for safety disappears (i.e., the potential field as a function of spacing $U(s_{ij})$ becomes a spike). Moreover, Newton's third law of motion holds if drivers respond to their surroundings isotropically. Furthermore, isotropic response, together with a hard collision, gives rise to the laws of momentum and energy conservation. Therefore, the phenomenology represents a special form of Newton's laws in a social setting (i.e., a transportation system involving human drivers). With its interpretation of the mean free path (i.e., desired car-following distance s_{ij}^*) and a molecular collision (i.e., action at a distance between vehicles), the phenomenology allows the application of other physical principles (such as kinetic theory) to further understand transportation systems as an ensemble.

21.6 SUMMARY

Involving both physical objects (e.g., vehicles) and living entities (e.g., drivers), a transportation system shares many commonalities with social and physical systems. The social side of the transportation system has long been recognized, as evidenced by applications of social rules and human factors in microscopic traffic flow modeling such as car following, lane changing, and route choice. In contrast, the physical side of the system has yet to receive proper attention. The transportation system does, however, exhibit many physical properties, which provides part of motivation for the proposed field theory of traffic flow.

To pave the foundation for the field theory, some physical phenomena in traffic flow were analyzed in relation to its social properties, in particular

motivations for drivers' decisions observed from their driving experiences. These phenomena, including those of mechanics, electromagnetics, waves, and statistical mechanics, strongly suggest that it is meaningful to integrate both physical and social principles into the modeling of traffic flow.

With the above understanding, the field theory was progressively formulated on the basis of a series of postulates, two of which are physical and two are social. The first postulate (physical) assumes that a roadway is a physical field in which a vehicle is subject to a roadway gravity and also a resistance due to the driver's willingness to observe traffic rules (e.g., speed limits). The second postulate (social) accounts for the driver's directional responsiveness to his or her surroundings. The third postulate (physical) imposes an action at a distance between two neighboring vehicles, and such an interaction is mediated by a potential field which is perceived as the danger of a collision. The distinction between a field perceived by a driver and a physical field is that the former impinges from the inside of the driver through motivation as opposed to through external compulsion. The fourth postulate (social) interprets driving strategy as a social rule—that is, a driver always tries to achieve gains (e.g., mobility and safety) and avoid losses (e.g., collisions and violation of traffic rules). From combination of the above postulates, the field theory was generically formulated as a system of equations governing the motion of a vehicle on a roadway in relation to other vehicles.

PROBLEMS

1. Use the field theory to explain the following phenomena:
 - a. A vehicle begins to accelerate after an emergency stop on the hard shoulder.
 - b. The vehicle gradually settles at its desired speed,
 - c. After a momentary speeding, the driver begins to speed up to the desired speed.
 - d. The driver applies the brakes when approaching a slow vehicle.
 - e. The driver adopts the leading vehicle's speed by following the leader.
 - f. The driver brakes again because a third vehicle cuts in between two vehicles in car-following mode.
 - g. The vehicle being cut off from the leading vehicle changes lane to seek speed gains.

2. Elaborate your strategies on how to capture the effect of the following traffic control devices with use of the field theory:
 - a. A signal
 - b. A STOP sign
 - c. A speed limit sign
3. Use the field theory to explain a rear-end accident.
4. Comment on how realistic the field theory is. Is it feasible to find evidence to prove or disprove the existence of a field?

CHAPTER 22

Longitudinal Control Model

A simple yet efficient traffic flow model, in particular one that describes vehicle longitudinal operational control and further characterizes the traffic flow fundamental diagram, is always desirable. Though many models have been proposed in the past, with each having its merits, research in this area is far from conclusive. This chapter introduces a new model—that is, the longitudinal control model (LCM)—with a unique set of properties to the arsenal. The model is suited for a variety of transportation applications, among which a concrete example is provided.¹

22.1 INTRODUCTION

A simple yet efficient traffic flow model, in particular one that describes vehicle longitudinal operational control and further characterizes the traffic flow fundamental diagram, is always desirable. For example, researchers can use such a model to study traffic flow phenomena, system analysts need the model to predict system utilization and congestion, accident investigators find the model handy to reconstruct accidents, software developers may implement the model to enable computerized simulation, and practitioners can devise strategies to improve traffic flow using such a simulation package.

Research has resulted in many traffic flow models, including microscopic car-following models and macroscopic steady-state models, each of which has its merits and is applicable in a certain context with varying constraints. An overview of these historical efforts will be provided in [Section 22.6](#). Nevertheless, research on traffic flow modeling is far from conclusive, and there is a constant quest for better models. Joining such a journey, this chapter presents a new model, the LCM, as a result of modeling from a combined perspective of physics and human factors ([Section 22.2](#)). The model seems to possess a unique set of properties:

- The model is *physically meaningful* because it captures the essentials of longitudinal vehicle control and motion on roadways with the presence of other vehicles ([Subsection 22.2.1](#))

¹ This chapter is reproduced from [113].

- The model is *simple* because it uses one equation to handle all driving situations in the longitudinal direction (Equation 22.2), and this microscopic equation aggregates to a steady-state macroscopic equivalent that characterizes the traffic stream in the entire density range (Equation 22.5).
- The model is *flexible* because the microscopic equation provides the mechanism to admit different safety rules that govern vehicle driving (Section 22.2.1), and the macroscopic equation has the flexibility to fit empirical traffic flow data from a variety of sources which exhibit varying flow-density relationships, including an reverse-lambda type (Section 22.3.2 and Figures 22.3–22.8).
- The model is *consistent* because the microscopic equation aggregates to its macroscopic equivalent, so the microscopic-macroscopic coupling is well defined (Section 22.2.2). As a result, traffic flow modeling and simulation based on the microscopic model aggregates to predictable macroscopic behavior (Section 22.5; see how the results of the microscopic and macroscopic approaches match).
- The model is *valid* as verified by field observations from a variety of locations (Section 22.4), and the model is realistic as demonstrated in an example application (Section 22.5).

The unique set of properties possessed by the LCM lend it to various transportation applications, including those mentioned above. An example of such applications is described in Section 22.5, where the LCM is applied to analyze traffic congestion macroscopically and microscopically. Research findings are summed up in Section 22.7.

22.2 THE LCM

Vehicle operational control in the longitudinal direction concerns a driver's response in terms of acceleration and deceleration on a highway without the worrying about steering, including lane changing. Rather than car following as it is conventionally termed, vehicle longitudinal control involves more driving regimes than simply car following (e.g., free flow, approaching, stopping). A field theory was proposed in Refs. [114, 115], and represents the environment (e.g., the roadway and other vehicles) perceived by a driver with ID i as an overall field U_i . As such, the driver is subject to forces as a result of the field. These forces, which impinge upon the driver's mentality, are the driving force G_i , roadway resistance R_i , and vehicle interaction F_i^j with the leading vehicle j (see Figure 22.1). Hence, the driver's response is

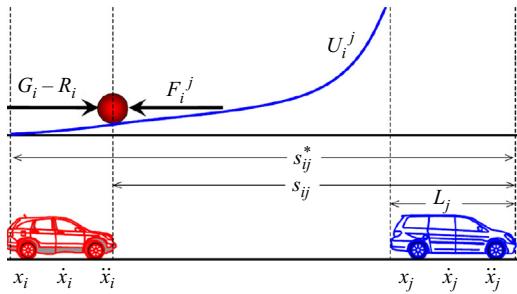


Figure 22.1 Forces acting on a vehicle.

the result of the net force $\sum F_i$ acting on the vehicle according to Newton's second law of motion:

$$\sum F_i = G_i - R_i - F_i^j. \quad (22.1)$$

22.2.1 Microscopic Model

If the functional forms of the terms in Equation 22.1 are carefully chosen (mainly by experimentation with empirical data), a special case called the LCM can be explicitly derived from Equation 22.1 as

$$\ddot{x}_i(t + \tau_i) = A_i \left[1 - \left(\frac{\dot{x}_i(t)}{v_i} \right) - e^{1 - \frac{s_{ij}(t)}{s_{ij}^*(t)}} \right], \quad (22.2)$$

where $\ddot{x}_i(t + \tau_i)$ is the operational control (acceleration or deceleration) of driver i executed after a perception-reaction time τ_i from the current moment t . A_i is the maximum acceleration desired by driver i when starting from standing still, \dot{x}_i is vehicle i 's speed, v_i is driver i 's desired speed, s_{ij} is the actual spacing between vehicle i and its leading vehicle j , and s_{ij}^* is the desired value of s_{ij} .

No further motivation for this special case is provided other than the following claims: (1) it takes a simple functional form that involves physically meaningful parameters but not arbitrary coefficients (see this and the next section), (2) it makes physical and empirical sense (see this section and Section 22.4), (3) it provides a sound microscopic basis for aggregated behavior—that is, traffic stream modeling (see the remainder of this section and Section 22.4)—and (4) it is simple and easy to apply (see Section 22.5).

The determination of the desired spacing $s_{ij}^*(t)$ admits safety rules. Basically, any safety rule that relates spacing to the driver's speed choice can

be inserted here. Of particular interest is an algorithm for the desired spacing that allows vehicle i to stop behind its leading vehicle j after a perception-reaction time τ_i and a deceleration process (at rate $b_i > 0$ which driver i believes that he or she is capable of applying in an emergency) should the leading vehicle j apply an emergency brake (at rate $B_j > 0$). After some math, the desired spacing can be determined as

$$s_{ij}^*(t) = \frac{\dot{x}_i^2(t)}{2b_i} - \frac{\dot{x}_j^2(t)}{2B_j} + \dot{x}_i\tau_i + l_j, \quad (22.3)$$

where $s_{ij}^* \geq l_j$ and l_j is vehicle j 's effective length (i.e., actual vehicle length plus some buffer spaces at both ends). Note that the term $\frac{\dot{x}_i^2(t)}{2b_i} - \frac{\dot{x}_j^2(t)}{2B_j}$ represents the degree of aggressiveness that driver i chooses. For example, when the two vehicles travel at the same speed, this term becomes $\gamma_i \dot{x}_i^2$, where

$$\gamma_i = \frac{1}{2} \left(\frac{1}{b_i} - \frac{1}{B_j} \right), \quad (22.4)$$

where B_j represents driver i 's estimate of the emergency deceleration which is most likely to be applied by driver j , while b_i is the deceleration which driver i believes that he or she is capable of applying in an emergency. It may be that b_i is greater than B_j in magnitude, which translates to the willingness (or aggressive characteristic) of driver i to take the risk of tailgating.

It is necessary to point out that though both B_j and b_i carry a sense of “emergency,” the model itself (i.e., Equations 22.2 and 22.3) is meant to describe all situations, including both “emergency” and “normal” operations. Or put it in another way, the LCM models a driver’s operational control \ddot{x}_i over a wide range on the basis of the interaction of a set of parameters, some of which concern the driver’s emergency responses—for example, B_j and b_i . This modeling philosophy echoes the “complete” car-following model described in Ref. [116, p. 158].

22.2.2 Macroscopic Model

Under steady-state conditions, vehicles in the traffic behave uniformly, and thus their identities can be dropped. Therefore, the microscopic LCM (Equations 22.2 and 22.3) can be aggregated to its macroscopic equivalent (traffic stream model):

$$v = v_f (1 - e^{1 - \frac{k^*}{k}}), \quad (22.5)$$

where v is traffic space mean speed, v_f is free-flow speed, k is traffic density, and k^* takes the following form:

$$k^* = \frac{1}{s^*} = \frac{1}{\gamma v^2 + \tau v + l}, \quad (22.6)$$

where γ and τ denote the aggressiveness and average response time, respectively, that characterize the driver population, and l denotes the average effective vehicle length. Equivalently, the macroscopic LCM can be expressed as

$$k = \frac{1}{s} = \frac{1}{(\gamma v^2 + \tau v + l)[1 - \ln(1 - \frac{v}{v_f})]}. \quad (22.7)$$

22.3 MODEL PROPERTIES

The LCM features a set of appealing properties that make the model unique. Firstly, it is a one-equation model that applies to a wide range of situations. More specifically, the microscopic LCM captures not only the car-following regime, but also other regimes, such as starting up, free flow, approaching, cutting off, and stopping (see Ref. [117] for more details). The macroscopic LCM applies to the entire range of density and speed without the need to identify break points.

Secondly, the LCM makes physical sense since it is rooted in basic principles (such as field theory and Newton's second law of motion). In addition, the LCM employs a set of model parameters that are not only physically meaningful but also easy to calibrate. For example, the microscopic LCM involves desired speed v_i , perception-reaction time τ_i , desired maximum acceleration when starting from standing still A_i , the deceleration which driver i believes that he or she is capable of applying in an emergency b_i , emergency deceleration B_j by driver j in front, and effective vehicle length l_j . The macroscopic LCM includes aggregated parameters, including free-flow speed v_f , aggressiveness γ , average response time τ , and effective vehicle length l . Data to calibrate the above parameters are either readily available in publications (such as *Motor Trend* and human factors study reports) or can be sampled in the field with reasonable efforts.

Lastly, the LCM is a consistent modeling approach—that is, the macroscopic LCM is derived from its microscopic counterpart when aggregated over vehicles and time. Such microscopic-macroscopic consistency not

only provides macroscopic modeling with a microscopic basis, but also ensures that microscopic modeling aggregates to a predictable macroscopic behavior.

More properties are discussed in the following subsections.

22.3.1 Boundary Conditions

The macroscopic LCM has two clearly defined boundary conditions. When density approaches zero ($k \rightarrow 0$), traffic speed approaches the free-flow speed ($v \rightarrow v_f$); when density approaches the jam density ($k \rightarrow k_j = 1/l$), traffic speed approaches zero ($v \rightarrow 0$) (see [Figure 22.9](#)).

One can determine the kinematic wave speed at jam density ω_j by finding the first derivative of flow q with respect to density k and evaluating the result at $k = k_j$. Hence,

$$q = kv = \frac{v}{(\gamma v^2 + \tau v + l)[1 - \ln(1 - \frac{v}{v_f})]}. \quad (22.8)$$

After some math,

$$\frac{dq}{dk} = v - \frac{s}{s'} = v - \frac{(\gamma v^2 + \tau v + l)[1 - \ln(1 - \frac{v}{v_f})]}{(2\gamma v + \tau)[1 - \ln(1 - \frac{v}{v_f})] + (\gamma v^2 + \tau v + l) \left(\frac{1}{v_f - v}\right)}. \quad (22.9)$$

Therefore, ω_j can be evaluated as

$$\omega_j = \left. \frac{dq}{dk} \right|_{k=k_j, v=0} = -\frac{l}{\tau + \frac{l}{v_f}}. \quad (22.10)$$

We can find capacity q_m by first setting Equation [22.9](#) to zero to solve for optimal speed v_m or optimal density k_m and then plugging v_m or k_m into Equation [22.8](#) to calculate q_m . However, it appears that an analytical solution of (q_m, k_m, v_m) is not easy to find, and this is a limitation of the LCM. Fortunately, the problem can be easily addressed numerically.

On another note, the spacing-speed relationship is

$$s = (\gamma v^2 + \tau v + l) \left[1 - \ln \left(1 - \frac{v}{v_f} \right) \right]. \quad (22.11)$$

We can determine the slope of the speed-spacing relationship when traffic is jammed by finding the first derivative of $v = f(s)$ with respect to spacing s and evaluating the result at $s = l$ and $v = 0$:

$$\begin{aligned} \left. \frac{dv}{ds} \right|_{s=l, v=0} &= \left. \frac{1}{(2\gamma v + \tau)[1 - \ln(1 - \frac{v}{v_f})] + (\gamma v^2 + \tau v + l) \left(\frac{1}{v_f - v} \right)} \right|_{s=l, v=0} \\ &= \left. \frac{1}{\tau + \frac{l}{v_f}} \right. . \end{aligned} \quad (22.12)$$

22.3.2 Model Flexibility

The macroscopic LCM employs four parameters that allow sufficient flexibility to fit data from a wide range of facilities (see the following section for details). As originally noted in Ref. [38] and later in Refs. [118, 119] concavity is a desirable property of the flow-density relationship. This property is empirically evident in field observations from most highway facilities, especially in outer lanes, and the shape of the flow-density relationship looks like a skewed parabola. In addition, some researchers [28, 31, 118–120] have recognized the attractiveness of having a triangular flow-density relationship. Moreover, an reverse-lambda shape was reported in Refs. [121, 122], most likely in the inner lane of freeway facilities. Therefore, a desirable property of a traffic stream model is its flexibility to represent a variety of flow-density shapes ranging from skewed parabola to triangular to reverse lambda.

The shape of the LCM is related to the second derivative of flow with respect to density:

$$\frac{d^2q}{dk^2} = -\frac{s^3 s''}{s'^3}, \quad (22.13)$$

where

$$s' = \frac{ds}{dv} = (2\gamma v + \tau) \left[1 - \ln \left(1 - \frac{v}{v_f} \right) \right] + (\gamma v^2 + \tau v + l) \left(\frac{1}{v_f - v} \right) \quad (22.14)$$

and

$$s'' = \frac{d^2s}{dv^2} = 2\gamma \left[1 - \ln \left(1 - \frac{v}{v_f} \right) \right] + \frac{4\gamma v + 2\tau}{v_f - v} + \frac{\gamma v^2 + \tau v + l}{(v_f - v)^2}. \quad (22.15)$$

Note that s is always positive, so the shape of the flow-density relationship is determined by the signs of s' and s'' . If s' and s'' are both positive,

d^2q/dk^2 is negative and the shape of the flow-density relationship is concave. Otherwise, the flow-density relationship may consist of a combination of concave, straight, and convex sections. In particular, it is possible to obtain an almost triangular shape and even an reverse-lambda shape under certain combinations of parameters v_f , τ , γ , and l , among which γ plays a critical role in controlling the shape of the flow-density relationship. For example, when the driver population is not aggressive—that is, $\gamma \geq 0$ —a concave flow-density relationship results; a moderately aggressive driver population may give rise to an almost triangular shape, and an aggressive driver population could lead to an reverse-lambda flow-density relationship.

The above discussion is further illustrated in Figure 22.2, where a family of fundamental diagrams are generated from the macroscopic LCM with the following parameters: $v_f = 30 \text{ m/s}$, $k_j = 0.2$ vehicles per meter, $\tau = 1 \text{ s}$, and aggressiveness γ ranging from 0 to $-0.03 \text{ s}^2/\text{m}$. In the flow-density plot, the lowest curve, exhibiting a skewed parabolic shape, is generated with $\gamma = 0$, the second highest curve, showing a nearly triangular shape, is generated with $\gamma = -0.027$, and the highest curve, which has an reverse-lambda shape, is generated with $\gamma = -0.030$. From the definition

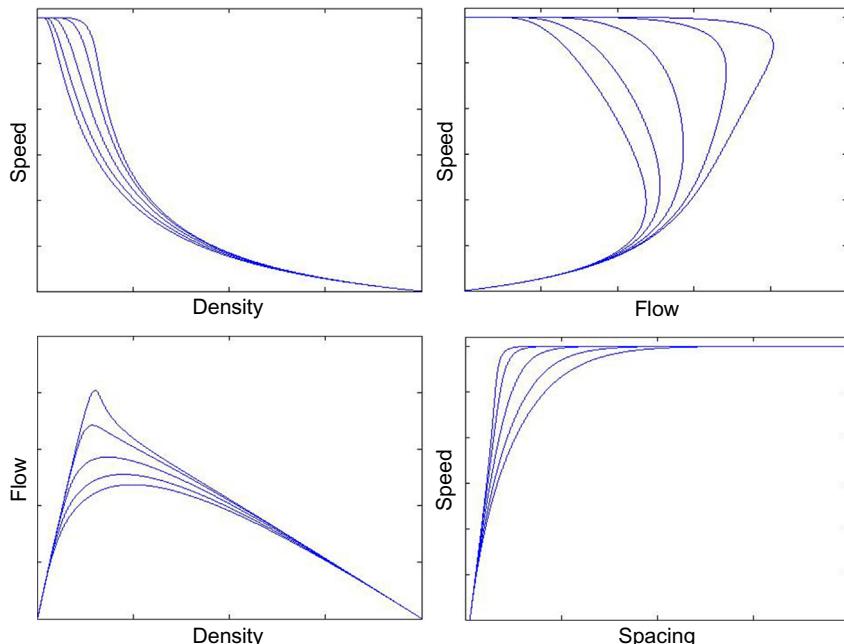


Figure 22.2 Family of curves generated from the LCM with different aggressiveness.

of aggressiveness in Equation 22.4, one recognizes that smaller values of γ correspond to more aggressive drivers, who are willing to accept shorter car-following distances. Therefore, the values of γ , the shape of the $q - k$ curves, and field observations are consistent. Further quantitative analysis of the effect of aggressiveness and its interaction with other model parameters warrants further research, and is not discussed here.

22.4 EMPIRICAL RESULTS

The LCM is tested by fitting the model to traffic flow data collected from a variety of facilities at different locations, including Atlanta (USA), Orlando (USA), Germany, California (USA), Toronto (Canada), and Amsterdam (Netherlands).

Figures 22.3–22.8 illustrate field data observed at these facilities with data “clouds” in the background labeled as “Empirical.” The fitted result of the LCM is illustrated as solid lines labeled as “LCM.” Also shown are the fitted results of other traffic stream models, including the Underwood model [11] (which employs two parameters) and the Newell model [58]

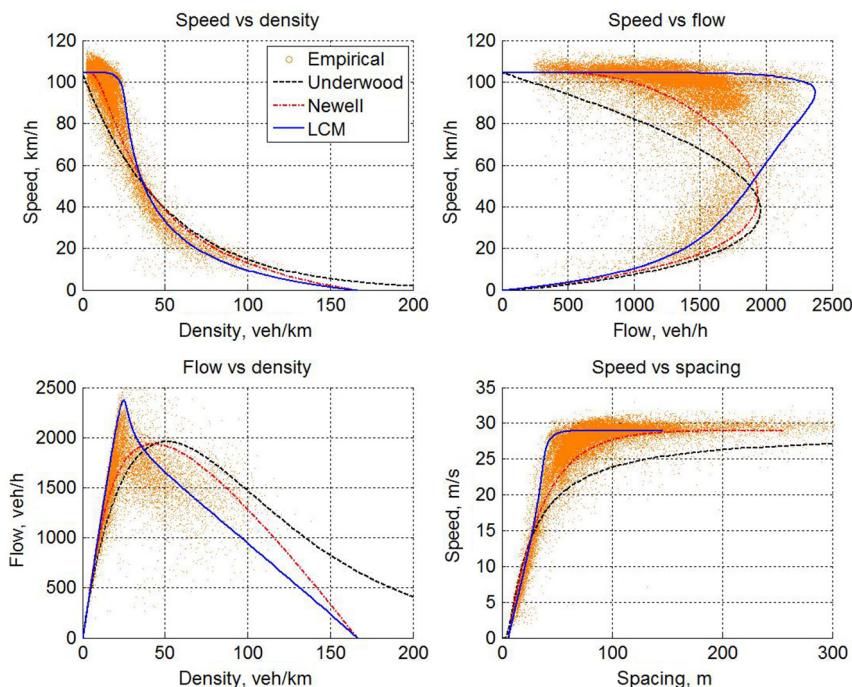


Figure 22.3 LCM fitted to GA400 data.

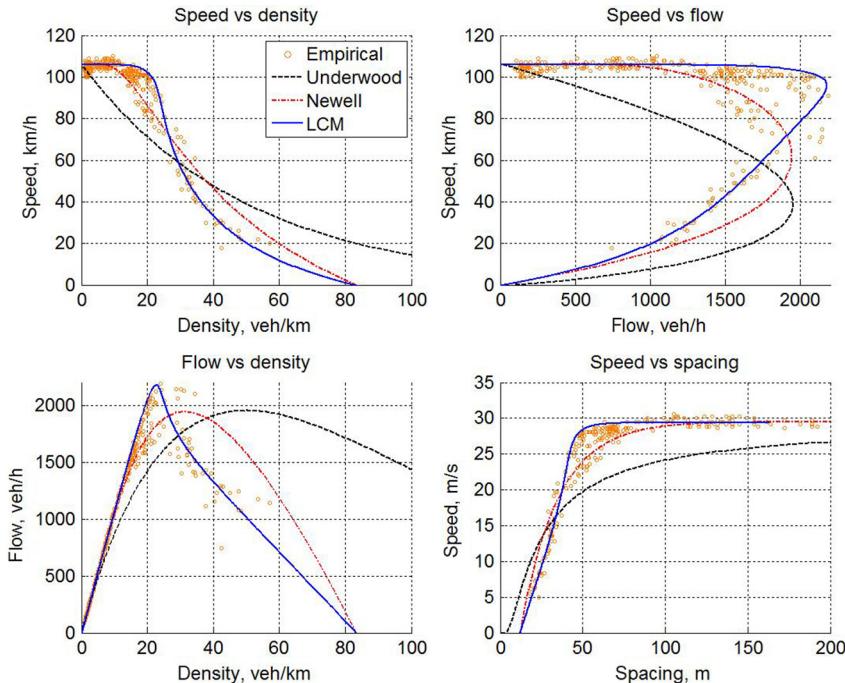


Figure 22.4 The LCM fitted to Interstate 4 data.

(three parameters). As such, the reader is able to visually compare the goodness of fit of two-, three-, and four-parameter models and examine how fit quality varies with the number of parameters. Consisting of four plots (namely, speed-density, speed-flow, flow-density, and speed-spacing), each figure illustrates the fundamental diagrams represented by empirical data and these models.

The empirical data in Figure 22.3 were collected on GA400, a toll road in Atlanta, Georgia, USA, at station 4001116. Consisting of 4787 observation points, the abundant field data reveal the relationships among flow, density, and speed by means of cloud density—that is, the intensity of data points. Meanwhile, the wide scatter of the data points seems to suggest that any deterministic, functional fit is merely a rough approximation, and a stochastic approach such as in Ref. [19] might be more statistically sound. By examining the cloud density, one is able to identify the trend of these relationships. For example, the flow-density relationship appears to have an reverse-lambda shape. Meanwhile, the speed-flow relationship features a \square shape with its “nose” leaning upward.

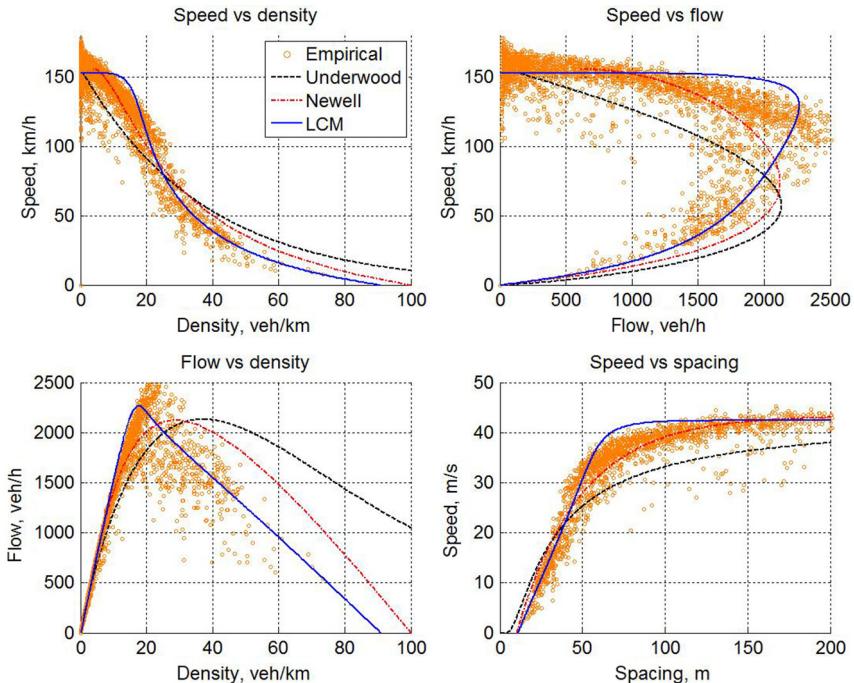


Figure 22.5 LCM fitted to autobahn data.

To fit the LCM to the empirical data, a two-level optimization procedure similar to that in Ref. [123] is adopted. First, each set of raw data is aggregated in order to reduce its size to a manageable level. When the data set is aggregated, its distribution with respect to density is obtained, and the entire density range is divided into intervals delimited by equally spaced quantiles. Then the data are aggregated by computation of an empirical mean (i.e., Emp mean) for each group consisting of the same number of consecutive observations. Next, the two-level optimization procedure is carried out. The inner loop searches for the minimum distance from each dot of “Emp mean” (ν_i , k_i , and q_i) to the LCM curve ($\hat{\nu}_i$, \hat{k}_i , and \hat{q}_i) normalized by (ν_f , k_j , and q_m) given a set of model parameters (ν_f , τ , γ , and l):

$$\min d_i = \sqrt{\left(\frac{\nu_i - \hat{\nu}_i}{\nu_f}\right)^2 + \left(\frac{k_i - \hat{k}_i}{k_j}\right)^2 + \left(\frac{q_i - \hat{q}_i}{q_m}\right)^2}. \quad (22.16)$$

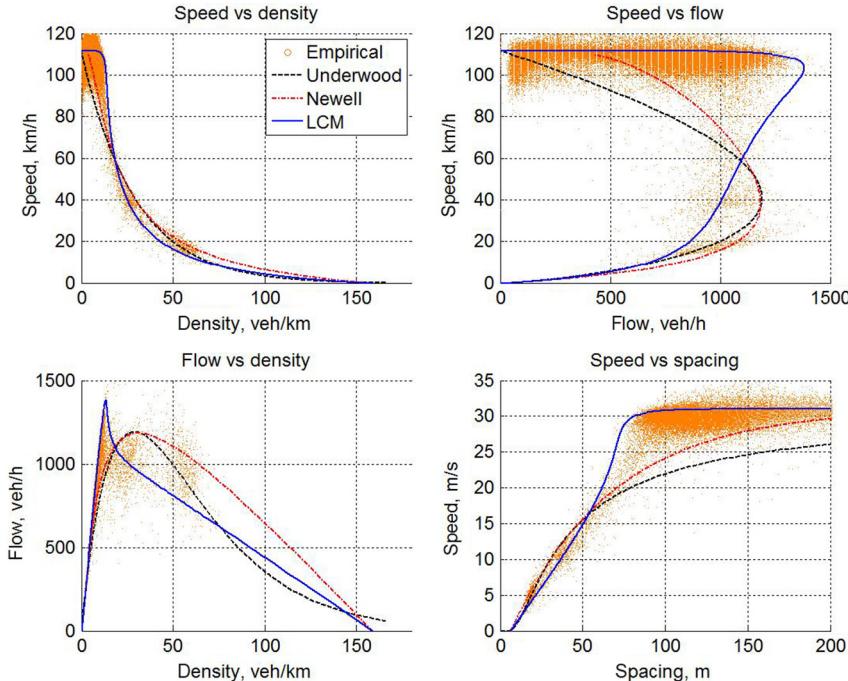


Figure 22.6 The LCM fitted to PeMS data.

Then, the outer loop searches for a set of optimized parameters that minimizes the total of minimized distances $D(\nu_f, \tau, \gamma, l)$:

$$\min D = \sum d_i \text{ subject to } \nu_f, \tau, \gamma, \text{ and } l. \quad (22.17)$$

Normally, this would end the fitting process. However, the optimized model does not always match the empirical capacity condition (q_m , k_m , and ν_m) since it consists only of a limited number of observations. If the capacity condition is also part of the fitting objective, one may need to tweak the optimized model, and this is typically done manually by visual inspection.

The fitting results are indicated in Tables 22.1 and 22.2. Table 22.1 compares the fitted capacity condition with the empirical capacity condition. The relative error of capacity is less than 5% and those of optimal density and speed are generally under 10%. Table 22.2 lists fitted parameters of the LCM. For example, the GA400 data set suggests a free-flow speed ν_f of 29 m/s (104.4 km/h), an effective vehicle length l of 6 m (or jam density $k_j = 167$ vehicles per kilometer), an average response time τ of 1.3 s, and

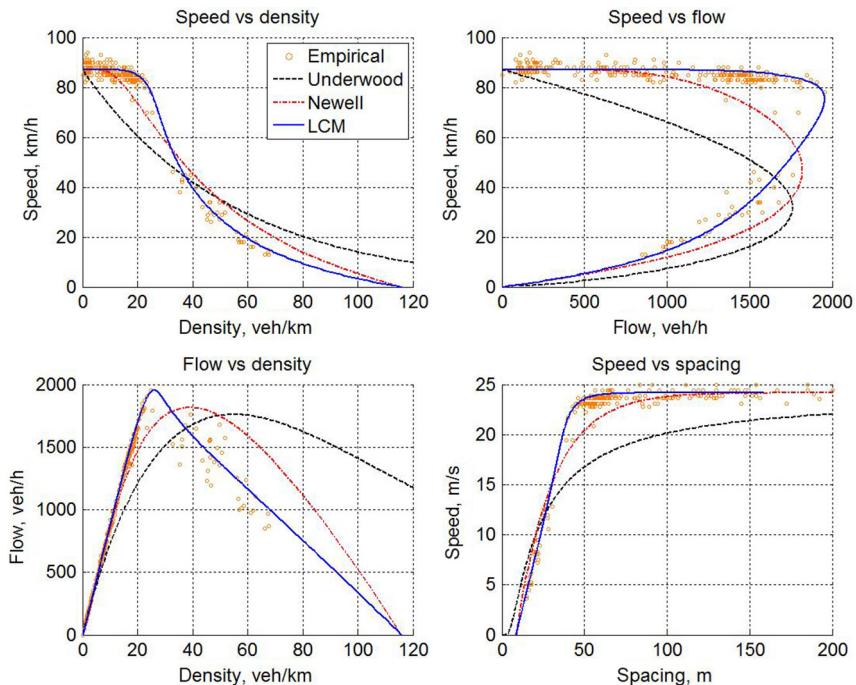


Figure 22.7 The LCM fitted to Highway 401 data.

aggressiveness γ of $-0.041 \text{ s}^2/\text{m}$. In addition, the kinematic wave speed at the jam condition ω_j is calculated with Equation 22.10. Though ω_j typically lies in a relatively narrow range between 15 and 25 km/h, outliers are observed in field data. For example, the CA/PeMS data set does not provide a clear clue to estimate ω_j , while the autobahn data set suggest an ω_j of 31.4 km/h or even higher.

We fitted two additional models to the data sets by matching empirical free-flow speed and capacity, and the results are presented in Table 22.2. It is apparent that the more parameters a model employs, the more flexible the model becomes, and hence the more likely it is to result in a good fit. In the speed-flow plot in Figure 22.3, the Underwood and Newell models are comparable in the congested regime (i.e., the lower portion of the graph), while in the free-flow regime (i.e., the upper portion of the graph) the Newell model outperforms the Underwood model since the Newell model is closer to the dense cloud. In contrast, the LCM (which employs four parameters) yields the best fit among the three models, as indicated by the close approximation of the LCM curve to the empirical

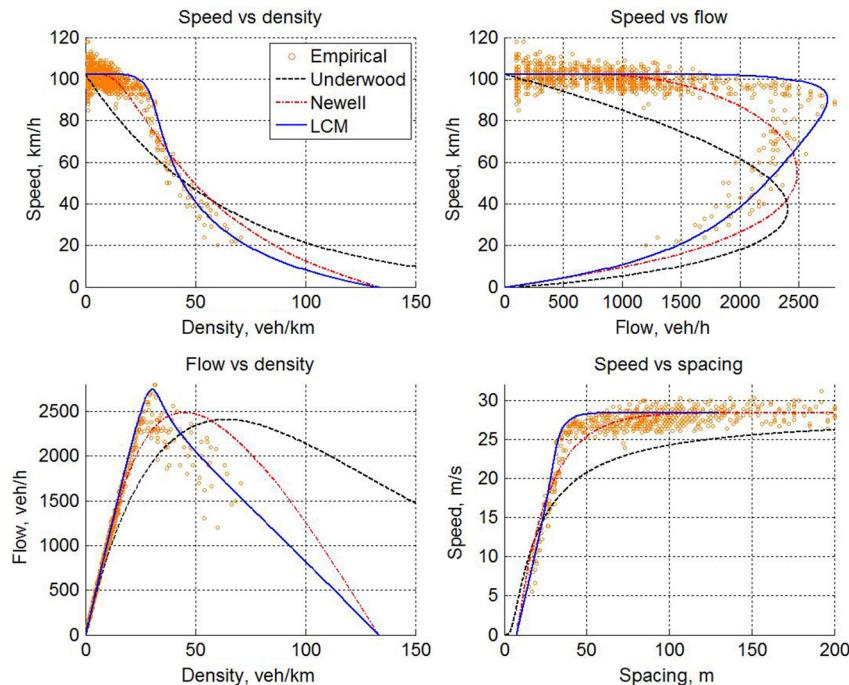


Figure 22.8 The LCM fitted to Amsterdam data.

data. More specifically, the LCM curve runs through the dense cloud in the free-flow regime and follows the trend nicely in the rest of the graph. In the flow-density plot, both the Underwood model and the Newell model peak later than do the empirical data. In the congested regime (i.e., the portion after the peak), both models exhibit a lack of fit, with the Newell model slightly better in terms of concavity and closeness to data points. In contrast, the LCM is superior on all accounts. Not only does it exhibit an reverse-lambda shape, it is also much closer to the empirical data. In addition, the curve peaks at the same location where the empirical data peak ($k_m = 25$ vehicles per kilometer). In the speed-density plot, the LCM appears to overfit the data when the density is very low. Except for this, the three models have their own relative merits since each appears to fit the empirical data reasonably well. The speed-spacing plot emphasizes the free-flow regime, which is the flat portion at the top of the graph. It appears that the Underwood model takes a long way to approach free-flow speed, while the Newell model and the LCM adapt to free-flow speed sooner.

Table 22.1 Parameters of the LCM as a result of fitting to various facility types

Data source		Empirical parameters					Capacity condition		
Location	Facility	No. of observations	v_f (m/s)	l (m)	τ (s)	γ (s^2/m)	q_m (vehicles/h)	k_m (vehicles/km)	v_m (km/h)
Atlanta	GA400	4787	29.5	4	1.46	-0.038	1883.8	22.0	85.8
Orlando	Interstate 4	288	24.2	8.6	1.09	-0.040	1795.5	22.1	81.4
Germany	Autobahn	3405	43.3	10	1.0	-0.018	2114.1	22.3	95.0
CA/PeMS	Freeway	2576	31	6.3	2.4	-0.060	1124.9	11.0	102.5
Toronto	Highway 401	286	29.5	12	0.8	-0.026	1945.7	21.8	89.2
Amsterdam	Ring road	1199	28.4	7.5	0.82	-0.026	2452.2	27.2	90.3

Table 22.2 Comparison of traffic stream models fitted to various facility types

Location	Model	Estimated parameters
Atlanta	Underwood	$v_f = 29.5 \text{ m/s}$, $k_m = 0.050 \text{ vehicles/m}$
	Newell	$v_f = 29.5 \text{ m/s}$, $l = 4.0 \text{ m}$, $\lambda = 0.81 \text{ 1/s}$
	LCM	$v_f = 29.5 \text{ m/s}$, $l = 4.0 \text{ m}$, $\tau = 1.46 \text{ s}$, $\gamma = -0.038 \text{ s}^2/\text{m}$
Orlando	Underwood	$v_f = 24.2 \text{ m/s}$, $k_m = 0.055 \text{ vehicles/m}$
	Newell	$v_f = 24.2 \text{ m/s}$, $l = 8.6 \text{ m}$, $\lambda = 1.09 \text{ 1/s}$
	LCM	$v_f = 24.2 \text{ m/s}$, $l = 8.6 \text{ m}$, $\tau = 1.09 \text{ s}$, $\gamma = -0.040 \text{ s}^2/\text{m}$
Germany	Underwood	$v_f = 43.3 \text{ m/s}$, $k_m = 0.037 \text{ vehicles/m}$
	Newell	$v_f = 43.3 \text{ m/s}$, $l = 10.0 \text{ m}$, $\lambda = 1.12 \text{ 1/s}$
	LCM	$v_f = 43.3 \text{ m/s}$, $l = 10.0 \text{ m}$, $\tau = 1.00 \text{ s}$, $\gamma = -0.018 \text{ s}^2/\text{m}$
CA/PeMS	Underwood	$v_f = 31.0 \text{ m/s}$, $k_m = 0.029 \text{ vehicles/m}$
	Newell	$v_f = 31.0 \text{ m/s}$, $l = 6.3 \text{ m}$, $\lambda = 0.50 \text{ 1/s}$
	LCM	$v_f = 31.0 \text{ m/s}$, $l = 6.3 \text{ m}$, $\tau = 2.40 \text{ s}$, $\gamma = -0.060 \text{ s}^2/\text{m}$
Toronto	Underwood	$v_f = 29.5 \text{ m/s}$, $k_m = 0.050 \text{ vehicles/m}$
	Newell	$v_f = 29.5 \text{ m/s}$, $l = 12.0 \text{ m}$, $\lambda = 1.3 \text{ 1/s}$
	LCM	$v_f = 29.5 \text{ m/s}$, $l = 12.0 \text{ m}$, $\tau = 0.80 \text{ s}$, $\gamma = -0.026 \text{ s}^2/\text{m}$
Amsterdam	Underwood	$v_f = 28.4 \text{ m/s}$, $k_m = 0.064 \text{ vehicles/m}$
	Newell	$v_f = 28.4 \text{ m/s}$, $l = 7.5 \text{ m}$, $\lambda = 1.5 \text{ 1/s}$
	LCM	$v_f = 28.4 \text{ m/s}$, $l = 7.5 \text{ m}$, $\tau = 0.82 \text{ s}$, $\gamma = -0.026 \text{ s}^2/\text{m}$

Unfortunately, the congested regime (the slope at the beginning portion of this graph) does not reveal much difference among the three models since they all cluster tightly together.

As shown in [Figure 22.4](#) and [Table 22.1](#), Interstate 4 data in Orlando, Florida, USA, feature a capacity q_m of 1953 vehicles per hour, which is achieved at an optimal density k_m of 24.9 vehicles per kilometer and optimal speed v_m of 78.4 km/h. What is striking in this set of data is that the free-flow regime in the speed-flow plot is almost flat and this condition is sustained almost up to capacity. This graph clearly differentiates the fitting quality of models with different numbers of parameters. More specifically, the two-parameter Underwood model exhibits the poorest fit since its upper branch (i.e., free-flow regime), nose (i.e., capacity), and lower branch (i.e., congested regime) are far from empirical observations. The three-parameter Newell model is better, as indicated by the closer fit of its upper branch, nose, and lower branch. The four-parameter LCM is superior in all aspects. For example, its upper branch is almost a flat line running through empirical data points, its nose leans upward and roughly coincides with empirically observed capacity, and its lower branch cuts evenly through empirical observations. Though there are discrepancies between the empirical data

and the fitted curve, no systematic overfitting or underfitting is observed in this graph. In the remaining three plots, the differences among the three models and their fit quality are consistent with those observed in the speed-flow plot.

In [Figure 22.5](#), the autobahn data collected from Germany exhibit an unusually high free-flow speed v_f of 42.4 m/s (or 152.6 km/h). Unlike the Interstate 4 data, which feature an almost constant free-flow speed v_f up to capacity, the traffic speed in the autobahn data gradually decreases in the free-flow regime, resulting in an optimal speed v_m of only about 60% of v_f as shown in the speed-flow plot. Unfortunately, the particular nature of this set of data poses a great challenge to any attempt to fit the data. In the speed-flow plot, one has difficulty to fit a model that meets the observed free-flow regime, the congested regime, and the capacity simultaneously, so a trade-off has to be made among the three portions. The LCM curve has been tweaked between free-flow and congested regimes while emphasizing the capacity. Though better than the Underwood and Newell models, the LCM still exhibits some discrepancies compared with the empirical data.

The PeMS data collected from California are plotted in [Figure 22.6](#). This set of data heavily emphasizes the free-flow regime (which is virtually a flat band in the speed-flow plot), with observations elsewhere sparsely scattered. In addition, a remarkable feature in the flow-density plot is the spike at capacity, which clearly indicates an reverse-lambda flow-density relationship. As expected, the LCM is able to be fitted to such a shape, and thus approximates the free-flow regime and the capacity condition very well. Since there are few observations in the congested regime, the model fit in this area appears to be quite arbitrary. In comparison, the LCM approximates the free-flow regime and the capacity condition the best, while the Underwood and Newell models are slanted and significantly underestimate the optimal speed v_m .

Though field observations on Highway 401 in Toronto do not have abundant data points, a trend is still clearly established in each plot in [Figure 22.7](#). Much like the results for the Interstate 4 data, there are clearly differences in capabilities among the models, with the two-parameter Underwood model being the poorest and the four-parameter model being the best. Notice that no systematic underfitting or overfitting is observed in the LCM curves. The same comments as above apply to ring road data in Amsterdam (see [Figure 22.8](#)).

22.5 APPLICATIONS

Since the LCM takes a simple mathematical form that involves physically meaningful parameters, the model can be easily applied to help investigate traffic phenomena at both the microscopic level and the macroscopic level. For illustrative purposes, a concrete example is provided below, in which a moving bottleneck is created by a sluggish truck. Microscopic modeling allows the LCM to generate profiles of vehicle motion so that the cause and effect of vehicles slowing down or speeding up can be analyzed in exhaustive detail; macroscopic modeling may employ the LCM to generate fundamental diagrams that help determine shock paths and develop graphical solutions. Since the LCM is consistent at the microscopic and macroscopic levels, the two sets of solutions not only agree with but also complement each other.

In addition, the LCM can be adopted by existing commercial simulation packages to improve their internal logic of car following, or it can serve as the basis of a new simulation package. Moreover, the LCM can be adopted in highway capacity and level of service analysis. For example, the conventional level of service analysis procedure involves the use of speed-flow curves to determine traffic speed; see Ref. [124] for the family of curves in EXHIBIT 23-3 and the set of approximating equations underneath. The macroscopic LCM can help make the analysis more effective by providing more realistic speed-density curves to facilitate analytical, numerical, and graphical solutions. Furthermore, the LCM can be adopted by transportation planners for use as the basis of a highway performance function which realistically estimates travel time (via traffic speed) as a function of traffic flow assigned to a route. The resultant travel time is the basis of driver route choice behavior, which in turn alters dynamic traffic assignment.

22.5.1 An Illustrative Example

A freeway segment contains an on-ramp (which is located 2000 m from an arbitrary reference point denoting the upstream end of the freeway) and an off-ramp 2000 m apart. The freeway was initially operating under condition A (flow 0.3333 vehicles per second or 1200 vehicles per hour, density 0.0111 vehicles per meter or 11.1 vehicles per kilometer, and speed 30 m/s or 108.0 km/h). At 2:30 p.m., a slow truck enters the freeway traveling at a speed of 5.56 m/s (20 km/h), which forces the traffic to operate under condition B (flow 0.3782 vehicles per second or 1361 vehicles per hour,

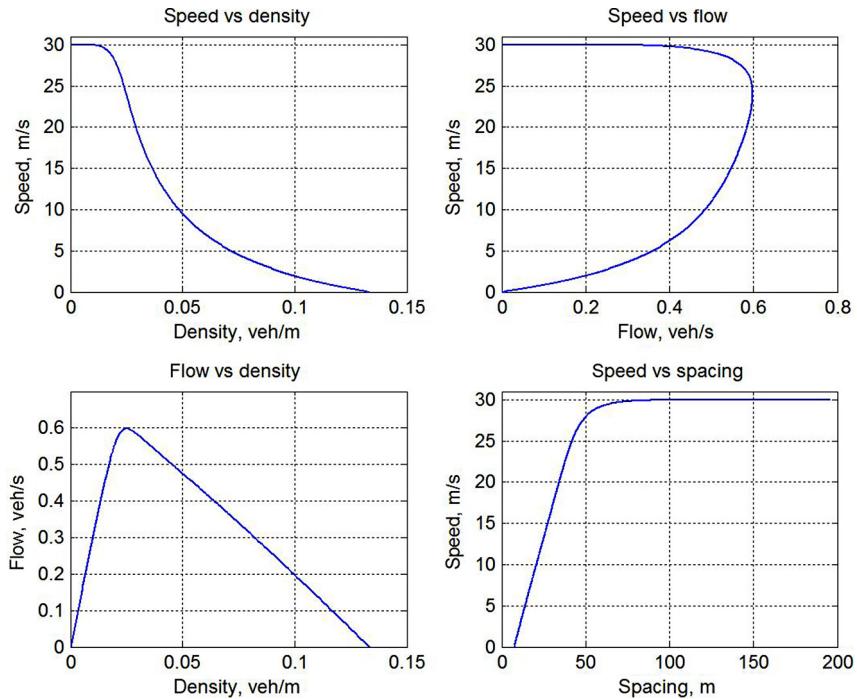


Figure 22.9 Fundamental diagram of the freeway generated from the LCM.

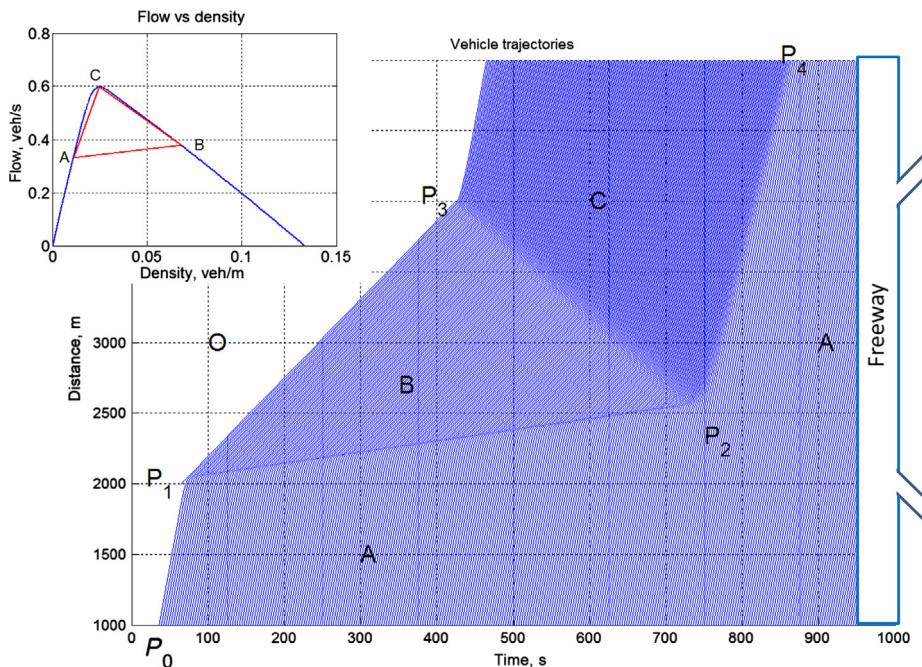
density 0.0681 vehicles per meter or 68.1 vehicles per /kilometer, and speed 5.56 m/s or 20 km/h). After a while, the truck turns off the freeway at the next exit. The impact on the traffic due to the slow truck is illustrated macroscopically in [Section 22.5.2](#) and microscopically in [Sections 22.5.3](#) and [22.5.4](#).

A fundamental diagram, which is illustrated in [Figure 22.9](#), is generated with the macroscopic LCM to characterize the freeway with the following parameters: free-flow speed $v_f = 30$ m/s, aggressiveness $\gamma = -0.028 \text{ s}^2/\text{m}$, average response time $\tau = 1$ s, and effective vehicle length $l = 7.5$ m. In addition, the above-mentioned traffic flow conditions, free-flow condition O, and capacity condition C are tabulated in [Table 22.3](#).

To illustrate the application of the LCM, the above problem is addressed in two approaches: a macroscopic graphical solution and a microscopic simulation solution. The microscopic simulation is conducted in deterministic and random fashions.

Table 22.3 Traffic flow conditions

Condition	Flow, q veh/s (veh/h)	Density k veh/m (veh/km)	Speed v m/s (km/h)
A	0.3333 (1200.0)	0.0111 (11.1)	30 (108.0)
B	0.3782 (1361.6)	0.0681 (68.1)	5.56 (20.0)
C	0.5983 (2154.0)	0.0249 (24.9)	24.03 (86.5)
O	0 (0)	0 (0)	30 (108.0)

**Figure 22.10** A moving bottleneck due to a slow truck; deterministic simulation.

22.5.2 Macroscopic Approach—Graphical Solution

The graphical solution to the problem involves finding shock paths that delineate time-space ($t-x$) regions of different traffic conditions. Figure 22.10 illustrates the time-space plane with the freeway overlaid on the right and a mini version of the flow-density plot in the top-left corner. The point when the slow truck enters the freeway (2:30 p.m.) roughly corresponds to $P_1(t_1 = 65, x_1 = 2000)$ on the time-space plane, while the point when the truck turns off the freeway is roughly $P_3(t_3 = 425, x_3 = 4000)$. Therefore, constrained by the truck, the $t-x$ region under P_1P_3 should contain traffic

condition B. On the other hand, the t - x regions before the truck enters the freeway and before congestion occurs (i.e., condition B) should have condition A. As such, there must be a shock path that delineates the two regions, and such a path should start at P_1 with a slope equal to shock wave speed U_{AB} , which can be determined according to Rankine–Hugoniot jump condition [125, 126]:

$$U_{AB} = \frac{q_B - q_A}{k_B - k_A} = \frac{0.3782 - 0.3333}{0.0681 - 0.0111} = 0.7877 \text{ m/s.} \quad (22.18)$$

Meanwhile, downstream of the off-ramp, congested traffic departs at capacity condition C, which corresponds to a t - x region that starts at P_3 and extends forward in time and space. Hence, a shock path forms between the region with condition C and the region with condition B. Such a shock path starts at P_3 and runs with a slope equal to shock wave speed U_{BC} :

$$U_{BC} = \frac{q_C - q_B}{k_C - k_B} = \frac{0.5983 - 0.3782}{0.0249 - 0.0681} = -5.0949 \text{ m/s.} \quad (22.19)$$

If the flow-density plot is properly scaled, one should be able to construct the above shock paths in the t - x plane. The two shock paths should eventually meet at point $P_2(t_2, x_2)$. Its location can be found by solving the following set of equations:

$$\begin{cases} x_2 - x_1 &= U_{AB} \times (t_2 - t_1), \\ x_2 - x_3 &= U_{BC} \times (t_2 - t_3), \\ (x_2 - x_1) + (x_3 - x_2) &= 2000. \end{cases} \quad (22.20)$$

After some math, P_2 is determined roughly at (716.8, 2513.4). After the two shock paths P_1P_2 and P_3P_2 meet, they both terminate and a new shock path forms which delineates regions with conditions C and A. The slope of the shock path should be equal to shock speed U_{AC} :

$$U_{AC} = \frac{q_C - q_A}{k_C - k_A} = \frac{0.5983 - 0.3333}{0.0249 - 0.0111} = 19.2029 \text{ m/s.} \quad (22.21)$$

As such, the shock path can be constructed as P_2P_4 . Lastly, the blank area in the t - x plane denotes a region with no traffic—that is, condition O.

22.5.3 Microscopic Approach—Deterministic Simulation

To double check the LCM and to verify if its macroscopic and microscopic solutions agree with each other reasonably, the microscopic LCM is implemented in MATLAB, a computational software package. As a manageable

starting point, the microscopic simulation is made deterministic with the following parameters: desired speed $v_i = 30 \text{ m/s}$, maximum acceleration $A_i = 4 \text{ m/s}^2$, emergency deceleration $B_i = 6 \text{ m/s}^2$, the deceleration which driver i believes that he or she is capable of applying in an emergency $b_i = 9 \text{ m/s}^2$, perception-reaction time $\tau_i = 1 \text{ s}$, and effective vehicle length $l_i = 7.5 \text{ m}$, where $i \in \{1, 2, 3, \dots, n\}$ are unique vehicle identifiers. Vehicles arrive at the upstream end of the freeway at a rate of one vehicle every 3 s, which corresponds to a flow of $q = 1200$ vehicles per hour. The simulation time increment is 1 s and the simulation duration is 1000 s.

[Figure 22.10](#) illustrates the simulation result in which vehicle trajectories are plotted in the t - x plane. The varying density of trajectories outlines a few regions with clearly visible boundaries. The motion or trajectory of the first vehicle is predetermined, while those of the remaining vehicles are determined by the LCM. The first vehicle enters the freeway at time $t = 65 \text{ s}$ (2:30 p.m.) after the simulation starts. This moment is calculated so that the second vehicle is about to arrive at the on-ramp at this particular moment. Hence, the second vehicle and vehicles thereafter have to adopt the speed of the truck, forming a congested region where traffic operates at condition B.

Upstream of this congested region B is a region where traffic arrives according to condition A. The interface of regions B and A, P_1P_2 , denotes a shock path in which vehicles in fast platoon A catch up with and join slow platoon B ahead. The situation continues, and the queue keeps growing until the truck turns off the freeway at $t = 425 \text{ s}$ into the simulation (2:36 p.m.). After that, vehicles at the head of the queue begin to accelerate according to the LCM—that is, traffic begins to discharge at capacity condition C. Therefore, the front of the queue shrinks, leaving a shock path P_3P_2 that separates region C from region B. Since the queue front shrinks faster than the growth of the queue tail, the former eventually catches up with the latter at P_2 , at which point both shock paths terminate, denoting the end of congestion. After the congestion disappears, the impact of the slow truck still remains because it leaves a capacity flow C in front, followed by a lighter and faster flow with condition A. Hence, the trace where faster vehicles in platoon A join platoon C denotes a new shock path P_2P_4 .

Comparison of the macroscopic graphical solution and the microscopic deterministic simulation reveals that they agree with each other very well, though the microscopic simulation contains much more information about the motion of each individual vehicle and the temporal-spatial formation and dissipation of congestion.

22.5.4 Microscopic Approach—Random Simulation

Since the microscopic approach allows one the luxury to account for randomness in drivers and traffic flow, the following simulation may replicate the originally posed problem more realistically. The randomness of the above example is set up as follows with the choice of distribution forms being rather arbitrary provided that they are convenient and reasonable:

- Traffic arrival follows a Poisson distribution, in which the headway between the arrival of consecutive vehicles is exponentially distributed with mean 3 s—that is, $h_i \sim \text{Exponential}(3)s$, which corresponds to a flow of 1200 vehicles per hour.
- The desired speed follows a normal distribution: $v_i \sim N(30, 2) \text{ m/s}$.
- The maximum acceleration follows a triangular distribution: $A_i \sim \text{Triangular}(3, 5, 4) \text{ m/s}^2$.
- Emergency deceleration $B_i \sim \text{Triangular}(5, 7, 6) \text{ m/s}^2$.
- The deceleration which driver i believes that he or she is capable of applying in an emergency $b_i \sim \text{Triangular}(8, 10, 9) \text{ m/s}^2$.
- Effective vehicle length $l_i \sim \text{Triangular}(5.5, 9.5, 7.5) \text{ m}$.

The result of one random simulation run is illustrated in [Figure 22.11](#), where the effect of randomness is clearly observable. Trajectories in region B seem to exhibit the least randomness because vehicles tend to behave uniformly under congestion. Trajectories in region C are somewhat random since the metering effect due to the congestion still remains. In contrast, region A appears to have the most randomness, not only because of the Poisson arrival pattern but also because of the random characteristics of drivers. Consequently, the shock path between regions B and C, P_3P_2 , remains almost unaltered, while there are some noticeable changes in shock path P_1P_2 . The first is the roughness of the shock path, and this is because vehicles in platoon A now join the tail of the queue in a random fashion. The second is that the path might not be a straight line. The beginning part of the shock path has a slope roughly equal to U_{AB} , while the rest has a slightly steeper slope (due to less intense arrival from the upstream part of the freeway during this period), resulting in the termination of congestion earlier than in the deterministic case (which is somewhere near P_2). This, in turn, causes the slope of the shock path between regions C and A to shift left. The slope of this shock path remains nearly the same since this scenario features a fast platoon that is caught by an even faster platoon.

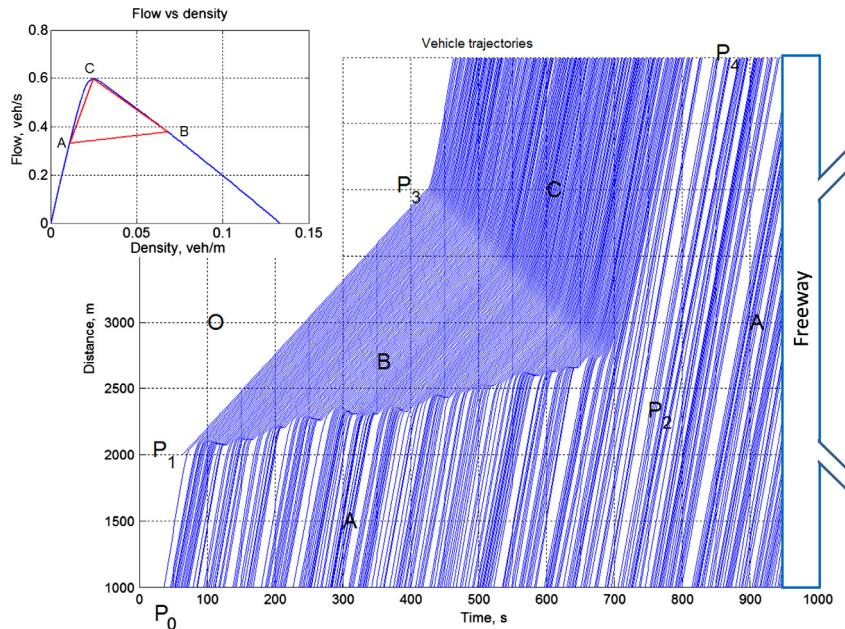


Figure 22.11 A moving bottleneck due to a slow truck; random simulation.

22.6 RELATED WORK

The microscopic LCM is a dynamic model which stipulates the desired motion (or acceleration) of a vehicle as the result of the overall field perceived by the driver. Other examples of dynamic models are General motors models [55, 56] and the intelligent driver model [60, 61, 116]. A dynamic model may reduce to a steady-state model when vehicle acceleration becomes zero. A steady-state model essentially represents a safety rule—that is, the driver's choice of speed as a result of allowing a safe car-following distance or vice versa. Examples of steady-state models include the Pipes model [52], the Forbes model [53, 54, 66], the Newell nonlinear car-following model [58], the Gipps car-following model [57], and the Van Aerde car-following model [62, 63]. Interested readers are referred to [127] for a detailed discussion of the relation among the LCM and other car-following models, including a unified diagram that summarizes such a relation.

The microscopic LCM incorporates a term called the desired spacing s_{ij}^* (Equation 22.2) which generally admits any safety rule and consequently any

steady-state model. However, Equation 22.3 instantiates s_{ij}^* in a quadratic form as a simplified version of the Gipps car-following model [57]. The result coincides with the speed-spacing relationship documented in the *Highway Capacity Manual* [128] and Chapter 4 of *Revised Monograph of Traffic Flow Theory* [3] as a result of 23 observational studies. The speed-spacing relationship incorporates three terms: a constant term representing effective vehicle length; a first-order term, which is the distance traveled during perception-reaction time τ ; a second-order term, which is the difference of the breaking distances of the following and leading vehicles, and which is interpreted as the degree of aggressiveness that the following driver desires to have. If one ignores the second-order term, the Pipes model [52] and equivalently the Forbes model [53, 54, 66] result.

The macroscopic model is a single-regime traffic stream (or equilibrium) model with four parameters. Also in the single-regime category, the Van Aerde model [62, 63] and the intelligent driver model [60, 61] employ four parameters, the Newell model [58] and the models of del Castillo [118, 120] have three parameters, and early traffic stream models such as those of Greenshields [9], Greenberg [10], Underwood [11], and Drake et al. [12] necessitate the use of only two parameters, though their flexibility and quality of fitting vary, as illustrated in Section 22.4.

22.7 SUMMARY

This chapter introduced a simple yet efficient traffic flow model, the LCM, which is a result of modeling from a combined perspective of physics and human factors. The LCM is formulated in two consistent forms: the microscopic model describes vehicle longitudinal operational control, and the macroscopic model characterizes steady-state traffic flow behavior and further the fundamental diagram.

The LCM was tested by fitting it to empirical data collected at a variety of facility types in different locations, including GA400 in Atlanta (USA), Interstate 4 in Orlando (USA), an autobahn in Germany, PeMS in California, Highway 401 in Toronto, and a ring road in Amsterdam. The wide scatter of these data sets suggests that any deterministic, functional fit is merely a rough approximation, and a stochastic approach might be more statistically sound. Test results support the claim that the LCM has sufficient flexibility to yield quality fits to these data sets, some of which even exhibit reverse-lambda flow-density relationships. Meanwhile, two more models are fitted to the same data sets in order to compare them

with LCM. These models include the two-parameter Underwood model and the three-parameter Newell model. The fitting results reveal that the more parameters a model employs, the more flexible the model becomes, and hence the more potential it has to achieve a good fit. Consistently, the Underwood model yields the poorest goodness of fit, while the Newell model represents an upgrade and the LCM maintains the best fit to empirical data.

The unique set of properties possessed by the LCM lend it to various transportation applications. For example, the LCM can be easily applied to help investigate traffic phenomena. An illustrative example was provided showing how to apply the LCM to the impact of a sluggish truck at both the microscopic level and the macroscopic level. Noticeably, the two sets of solutions agree with and complement each other owing to the consistency of the LCM. In addition, the LCM can be adopted by existing commercial simulation packages to improve their internal logic of car following, or perhaps serve as the basis of a new simulation package. Moreover, the LCM may help make highway capacity and level of service analysis more effective by providing more realistic speed-density curves to facilitate analytical, numerical, and graphical solutions. Furthermore, the LCM can assist in effective transportation planning by providing a better highway performance function that helps determine driver route choice behavior.

PROBLEMS

1. Vehicle i has just resumed motion after an emergency stop on the hard shoulder of a freeway. According to the microscopic form of the LCM, how long does it take for the vehicle to bring its speed up to 20 m/s? Assume the parameter values specified [Section 22.5.3](#) apply.
2. After some time, vehicle i is approaching a leading vehicle j , at which time vehicle i is traveling at 25 m/s, vehicle j is traveling at 20 m/s, and the spacing between the two vehicles is 50 m. Use the LCM and use parameter values specified above to calculate the following under this scenario:
 - a. Driver i 's desired spacing
 - b. Driver i 's control decision
 - c. When driver i executes the control decision
3. Building on the above scenario, at this time a third vehicle k traveling at 23 m/s on the adjacent lane cuts in halfway between vehicles i and j .

Assume the same set of parameters apply to vehicle k as well. Repeat the tasks in problem 2 for vehicles i and k .

4. For the scenario above, if the underlying desired spacing model is replaced by the Forbes model and the parameter values remain the same, how would your answers to the above problem change?
5. Assume that the parameters in the above problem apply and are uniform across all drivers and vehicles, and determine the corresponding macroscopic form of the LCM.
6. Use the above macroscopic form of the LCM to determine the following:
 - a. Kinematic wave speed at jam density
 - b. The slope of the speed-spacing relationship when traffic is jammed.
7. Find the capacity condition of the above macroscopic form of the LCM.
8. A unique feature of the macroscopic LCM is its ability to generate different shapes of flow-density curves by varying parameter values within the same functional model form. Use the above set of parameters as a starting point and tweak parameters of your choice to configure the following types of flow-density curves and indicate your set of parameter values for each curve:
 - a. A skewed parabolic shape
 - b. A triangular shape
 - c. An reverse-lambda shape

CHAPTER 23

The Unified Diagram

Using the field theory and the longitudinal control model presented in the previous two chapters as a framework, one can conveniently relate traffic flow models to each other. As such, a unified perspective can be cast on traffic flow modeling with bridges not only within but also between microscopic and macroscopic levels.¹

23.1 MOTIVATION

Half a century ago, Newell [58] proposed a nonlinear car-following model of the following form:

$$\dot{x}_i(t + \tau_i) = v_i \left(1 - e^{-\frac{\lambda_i}{v_i}(s_{ij}(t) - l_i)} \right), \quad (23.1)$$

where $\dot{x}_i(t)$ is the speed of the vehicle with ID i at time t , τ_i is driver i 's perception-reaction time, v_i is driver i 's desired speed, λ_i is a parameter associated with driver i (i.e., the slope of driver i 's speed-spacing curve evaluated at $\dot{x}_i = 0$), s_{ij} is the spacing between vehicle i and its leader bearing ID j , and l_i is the minimum value of s_{ij} . Newell acknowledged that “no motivation for this choice is proposed other than the claim that it has approximately the correct shape and is reasonably simple.”

It would be interesting to interpret the Newell model and furnish it with a possible motivation (this section). In doing so, we find that the interpretation gives rise to a broader picture that can be used to relate some existing traffic flow models to each other (Section 23.2). As such, a unified perspective can be cast on traffic flow modeling with bridges not only within but also between microscopic and macroscopic levels (Section 23.3).

Without further delay, the Newell model can be slightly rearranged as follows:

$$1 - \frac{\dot{x}_i(t + \tau_i)}{v_i} - e^{\frac{l_i - s_{ij}(t)}{v_i/\lambda_i}} = 0 \quad (23.2)$$

¹ This chapter is reproduced from Ref. [129].

The above equation is, in turn, a special case of the following equation when vehicle i 's acceleration \ddot{x}_i is zero:

$$\ddot{x}_i(t + \tau_i) = g_i \left[1 - \frac{\dot{x}_i(t)}{v_i} - e^{\frac{l_i - s_{ij}(t)}{v_i/\lambda_i}} \right], \quad (23.3)$$

where g_i is a positive, nonzero parameter associated with vehicle i . Equation (23.3) is a *dynamic* car-following model which describes the acceleration performance of vehicle i , whereas Equation (23.2) is a *steady-state* version of the dynamic model since the former describes the speed choice of driver i in the steady state—that is, when acceleration is not considered ($\ddot{x}_i = 0$).

Steady-state and dynamic car-following models are both widely applied and successful in microscopic traffic flow simulation. However, dynamic models appear to be more desirable in modeling driver operational control (e.g., car following) if the following two issues are concerned. The first pertains to human factors. Though a driver may have a speed choice in mind (e.g., “I wish to travel at 113 km/h (or 70 miles per hour)”), such a goal is achieved over time, during which time the driver’s operational control at each instant is based on acceleration (e.g., “I need to speed up or slow down in order to get to the target speed”), which naturally results from the driver’s operation on the gas and brake pedals. The second pertains to physics. The acceleration of an object can change abruptly, whereas its speed profile has to be smooth. For example, when a driver steps on the brake pedal and keeps the foot there to bring the vehicle to a stop, a deceleration is constantly applied until the vehicle stops, at which moment the deceleration suddenly disappears. In contrast, the speed profile of the vehicle has to be smooth—that is, starting from its initial speed and continuously decreasing to zero. As another example, when a subject vehicle is being cut off from its leader (because of a third vehicle squeezing in between), the sudden change of spacing may result in a steady-state model model to suggest an unattainable speed in response (at which point, an extra, *dynamic* constraint on limiting acceleration has to be introduced which exceeds the scope of steady-state modeling). In contrast, a dynamic model works directly on acceleration, and even though limiting acceleration may be involved, it is still within the scope of dynamic modeling.

We can further rearrange Equation (23.3) as follows by multiplying both sides by vehicle mass m_i :

$$m_i \ddot{x}_i(t + \tau_i) = m_i g_i - m_i g_i \frac{\dot{x}_i(t)}{v_i} - m_i g_i e^{\frac{l_i - s_{ij}(t)}{v_i/\lambda_i}}. \quad (23.4)$$

One immediately recognizes that the above equation takes the form of Newton's second law of motion:

$$\sum F_i = G_i - R_i - F_i^j, \quad (23.5)$$

where $\sum F_i = m_i \ddot{x}_i(t + \tau_i)$, $G_i = m_i g_i$, $R_i = G_i \frac{\dot{x}_i(t)}{v_i}$, and $F_i^j = G_i e^{\frac{l_j - s_{ij}(t)}{v_i/\lambda_i}}$. Therefore, Equation (23.4) can be interpreted as an application of Newton's second law of motion in driver operational control. The acceleration of a driver-vehicle unit i is the result of "forces" acting on the unit, and these forces can be further interpreted as follows. The term G_i functions as the driving force, which is analogous to the gravity and is determined as the product of vehicle mass m_i and the acceleration of roadway gravity g_i . The term R_i is like a resistance: the faster the vehicle travels, the greater the resistance is. In addition, the resistance balances the gravity when the driver's desired speed is achieved. The term F_i^j can be interpreted as a repelling (vehicle interaction) force from leading vehicle j depending on the spacing s_{ij} between the two vehicles. Since this is a noncontact force, it is an action at a distance as if it were mediated by a "field."

23.2 A BROADER PERSPECTIVE

Extending the above discussion, we find it appropriate to interpret the driver's operational control using the concept of a field. More specifically, the driving environment perceived by a driver can be represented as a field, in which objects (such as roadways and other vehicles) are each represented as a component field and their superposition represents the overall hazard that the subject driver tries to avoid. Hence, the aim of modeling of vehicle motion is to seek the least hazardous route by navigating through the field along its valley. A field theory of such a nature was introduced in Chapter 21. Only major results of the field theory are reproduced below for easy reference.

23.2.1 Overview of the Field Theory

The generic form of the field theory is

$$\begin{cases} m_i \ddot{x}_i(t + \tau_i) &= \gamma_i^0 [G_i(t) - R_i(t)] + \gamma (\alpha_i^j) \frac{\partial U_{i,x}}{\partial x}, \\ m_i \ddot{y}_i(t + \tau_i) &= -\gamma (\alpha_i^k) \frac{\partial U_{i,y}}{\partial y}. \end{cases} \quad (23.6)$$

Readers are referred to Chapter 21 for the derivation of the results and the notation presented in this subsection. A special case of the theory, which

is referred to as the longitudinal control model, can take the following form after some simplifications:

$$\ddot{x}_i(t + \tau_i) = g_i \left[1 - \left(\frac{\dot{x}_i(t)}{v_i} \right)^\delta - e^{\frac{s_{ij}(t)^* - s_{ij}(t)}{Z}} \right] \quad (23.7)$$

or

$$\ddot{x}_i(t + \tau_i) = g_i \left[1 - \left(\frac{\dot{x}_i(t)}{v_i} \right) - e^{\frac{s_{ij}(t)^* - s_{ij}(t)}{s_{ij}(t)^*}} \right] \quad (23.8)$$

if one chooses $\delta = 1$ and $Z = s_{ij}(t)^*$. The desired spacing $s_{ij}(t)^*$ is motivated by safety rules and can take many forms, of which two examples are provided:

$$s_{ij}^*(t) = \dot{x}_i(t)\tau_i + l_j, \quad (23.9)$$

$$s_{ij}^*(t) = \frac{\dot{x}_i^2(t)}{2b_i} + \dot{x}_i(t)\tau_i - \frac{\dot{x}_j^2(t)}{2B_j} + l_j. \quad (23.10)$$

Aggregating [Equation 23.8](#) over vehicles by assuming steady-state conditions yields the following equilibrium speed-density relationship:

$$v = v_f [1 - e^{1 - \frac{k^*}{k}}]. \quad (23.11)$$

A more specific form is

$$v = v_f \left[1 - e^{1 - \frac{1}{k(\gamma v^2 + \tau v + l)}} \right] \quad (23.12)$$

or

$$k = \frac{1}{(\gamma v^2 + \tau v + l) \left[1 - \ln \left(1 - \frac{v}{v_f} \right) \right]}. \quad (23.13)$$

23.2.2 Relating Microscopic Car-Following Models

Following the rationale in [Section 23.1](#), it turns out that the field theory can be used as a framework to relate traffic flow models to each other at both the microscopic level and the macroscopic level.

Newell Model

We return to the Newell nonlinear car-following model. Comparison of [Equations 23.1](#) and [23.7](#) reveals that the former results if one chooses to (1) apply the steady-state condition—that is, $\ddot{x}_i(t + \tau_i) = 0$; (2) set $Z = v_i/\lambda_i$; (3) let $s_{ij}^*(t) = l_i$; and (4) use $\dot{x}_i(t)$ as the response variable and apply a time delay τ_i .

Further, the physical meaning of parameter λ_i in Equation 23.1 is the tangent of the speed-spacing curve (Figure 1 in Newell's original paper) evaluated when the speed is zero. This parameter can be interpreted as the reciprocal of the perception-reaction time (i.e., $\lambda_i = 1/\tau_i$) as implied by Newell's Figure 1 and the numerical values in his Figure 2). In contrast, this tangent is evaluated as $1/(\tau + l/v_f)$ in the longitudinal control model. With this understanding, the vehicle interaction force F_i^j suggested by the Newell model can be interpreted as the negative exponential of the gap ($s_{ij}(t) - l_i$) between the subject vehicle i and its leader j scaled down by the distance ($v_i\tau_i$) traversed by vehicle i at the desired speed v_i during one perception-reaction time τ_i .

The field theory is related to other microscopic car-following models as follows.

Forbes Model

The Forbes model [53, 54, 66] is based on the following safety rule: the time gap between a vehicle and its leader should always be equal to or greater than the reaction time τ_i . This model can be admitted into the longitudinal control model (Equation 21.15) as a means to determine the *desired spacing* $s_{ij}^*(t)$, which is formulated in Equation 23.9.

General Motors Models

The family of General Motors models (GM models) [55, 56] is generically formulated in its fifth model (GM5):

$$\ddot{x}_i(t + \tau_i) = \alpha \frac{\dot{x}_i^m(t + \tau_i)[\dot{x}_j(t) - \dot{x}_i(t)]}{[x_j(t) - x_i(t)]^l}. \quad (23.14)$$

If one chooses $m = l = 1$, Equation 23.14 reduces to the fourth-generation GM model (GM4):

$$\ddot{x}_i(t + \tau_i) = \alpha \frac{\dot{x}_i(t + \tau_i)[\dot{x}_j(t) - \dot{x}_i(t)]}{[x_j(t) - x_i(t)]}, \quad (23.15)$$

where x_i , \dot{x}_i , \ddot{x}_i , and τ_i are the displacement, speed, acceleration, and perception-reaction time of the subject vehicle i , respectively; similar notation applies to its leader j ; α is a dimensionless coefficient. In relation to the field theory, GM4 considers only the vehicle interaction force F_i^j and ignores the unsatisfied desire for mobility ($G_i - R_i$) (see Chapter 21 for details). Rather than translating intrusion exponentially to vehicle interaction force as in the longitudinal control model (Equation 21.15), F_i^j

in GM4 mimics Coulomb's law in electrostatics. More specifically, GM4 views vehicle i as a particle which carries a moving coordinate with electric charge equivalent to its speed \dot{x}_i and vehicle j as another particle which moves relative to vehicle i with charge equivalent to their relative speed $[\dot{x}_j(t) - \dot{x}_i(t)]$. The magnitude of the interaction force is proportional to the product of the two charges and inversely proportional to their distance. According to [Equation 23.15](#), vehicle i is attracted to (or repelled by) vehicle j if the latter travels faster (or slower) than the former.

Gipps Model

The Gipps model [\[57\]](#) consists of a system of two inequalities with one governing the free-flow regime and the other governing the car-following regime.

The free-flow inequality reproduced below is a result of fitting empirical observations, and its function is to accelerate a vehicle from its initial speed asymptotically toward its desired speed without oscillation:

$$\dot{x}_i(t + \tau_i) = \dot{x}_i(t) + 2.5g_i\tau_i \left(1 - \frac{\dot{x}_i(t)}{v_i}\right) \sqrt{0.025 + \frac{\dot{x}_i(t)}{v_i}}. \quad (23.16)$$

We can rewrite the above equation in the following differential form after considering the time difference τ :

$$\ddot{x}_i(t + \tau_i) \approx \frac{\dot{x}_i(t + \tau_i) - \dot{x}_i(t)}{\tau_i} = g'_i \left(1 - \frac{\dot{x}_i(t)}{v_i}\right), \quad (23.17)$$

where $g'_i = 2.5g_i\sqrt{0.025 + \frac{\dot{x}_i(t)}{v_i}}$. Note that [Equation 23.17](#) is actually the *unsatisfied desire for mobility* term in [Equation 21.15](#) when the vehicle interaction term disappears.

The Gipps car-following model is derived from the following safety rule: at any moment, a driver i should leave sufficient distance behind the leader j such that driver i has enough room to respond and decelerate at a rate of $b_i > 0$ to a safe stop behind j should the leader apply an emergency brake ($B_j > 0$). The scenario is illustrated in [Figure 23.1](#) and the model reproduced below:

$$s_{ij}^*(t) \geq \frac{\dot{x}_i(t + \tau_i)^2}{2b_i} + \frac{\tau_i}{2}[\dot{x}_i(t) + \dot{x}_i(t + \tau_i)] + \dot{x}_i(t + \tau_i)\theta - \frac{\dot{x}_j^2}{2B_j} + L. \quad (23.18)$$

The astute reader has recognized that the above model describes the *desired spacing*, which follows exactly the same safety rule used to derive

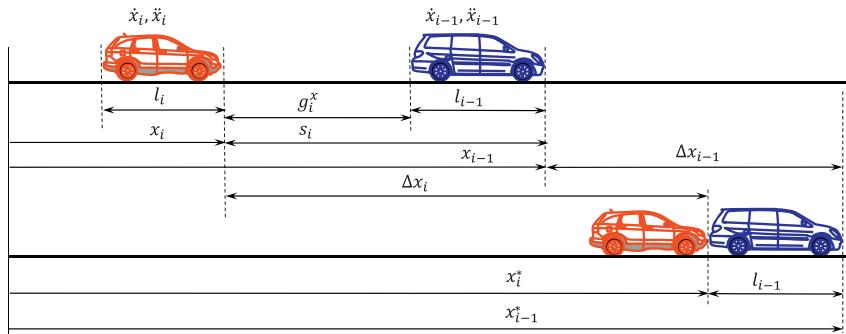


Figure 23.1 The Gipps model.

[Equation 23.10](#), which is slightly modified from and simpler than the above model. Of course, one could opt to use this model in place of [Equation 23.10](#) to apply the longitudinal control model.

Note that the Gipps model has been identified as “overly safe” because of the rather conservative safety rule and the additional safety margin $\dot{x}_i(t + \tau_i)\theta$. Consequently, excessive car-following distances result, and the model significantly underestimates highway capacity. In reality, though these safety measures make sense, drivers tend to use them as a good rule of thumb but frequently follow other vehicles closer than the desired spacing.

Intelligent Driver Model

The intelligent driver model (IDM) [60, 61] is expressed as a superposition of the follower i 's acceleration term and a deceleration term which depends on the desired spacing s_{ij}^* :

$$\ddot{x}_i(t + \tau_i) = g_i \left[1 - \left(\frac{\dot{x}_i}{v_i} \right)^\delta - \left(\frac{s_{ij}^*}{s_{ij}} \right)^2 \right], \quad (23.19)$$

where δ is the acceleration exponent, $s_{ij} = x_j - x_i$ is the spacing between vehicle i and its leader j , and the desired spacing s_{ij}^* is a function of speed \dot{x}_i and relative speed $(\dot{x}_i - \dot{x}_j)$: $s_{ij}^* = s_0 + s_1 \sqrt{\dot{x}_i/v_i} + T_i \dot{x}_i + \dot{x}_i [\dot{x}_i - \dot{x}_j]/[2\sqrt{g_i b_i}]$, where s_0 , s_1 , b_i , and T_i are parameters. Compared with Equation 21.15, the IDM strikingly resembles the longitudinal control model. From the perspective of the field theory, the IDM relates the interaction F_i^j between vehicle i and its leader j to the squared ratio of the desired spacing s_{ij}^* to the actual spacing s_{ij} . In addition, the IDM has its own safety rule to determine s_{ij}^* which is conveniently admittable to the longitudinal control model.

Van Aerde Model

The Van Aerde car-following model [62, 63] combines the Pipes model [52] and the Greenshields model [9] into a single equation:

$$s_{ij} = c_1 + c_3 \dot{x}_i + c_2 / (\nu_f - \dot{x}_i), \quad (23.20)$$

where $c_1 = \nu_f(2\nu_m - \nu_f)/(k_j \nu_m^2)$, $c_2 = \nu_f(\nu_f - \nu_m)^2/(k_j \nu_m^2)$, $c_3 = 1/q_m - \nu_f/(k_j \nu_m^2)$, where ν_f is the free-flow speed of the roadway facility, k_j is the jam density, and ν_m is the optimal speed occurring at capacity q_m .

The Van Aerde model constitutes yet another safety rule which can be related to the longitudinal control model as the desired spacing s_{ij}^* .

CARSIM Model

The CARSIM model [65] consists of a set of acceleration algorithms (reproduced below to be consistent in notation):

- A1:** Vehicle i is moving but has not yet reached its desired speed ν_i . Depending on vehicle i 's initial speed and the urgency of the task, the acceleration rate is found by entering the data in Tables 1 and 2 in Ref. [65].
- A2:** Vehicle i has reached its desired speed ν_i . No specific algorithm is provided except that the driver will try to reach ν_i as fast as possible while satisfying all safety and operational constraints.
- A3:** Vehicle i was stopped and has to start from standstill. A maximum acceleration rate is applied constrained by a noncollision requirement after a response delay τ_i .
- A4:** Vehicle i is in car-following mode with its leader j . A_4 is determined by satisfying the following safety rule: vehicle i should leave a nonnegative gap ($s_{ij} - l_j \geq 0$) from leader j should vehicle i be advanced one time increment Δt : $s_{ij}(t) = x_j(t) - x_i(t + \Delta t) \geq l_j$, where $x_i(t + \Delta t) = x_i(t) + \dot{x}_i \Delta t - 0.5 A_4 \Delta t^2$ and the other variables are as defined before.
- A5:** Vehicle i in car-following mode is subject to a noncollision constraint which is reinforced by considering the desired spacing: $s_{ij}^*(t) = x_j(t) - x_i(t + \Delta t) \geq \max\{\dot{x}_i(t + \Delta t)\tau_i + l_j; \text{or } \dot{x}_i(t + \Delta t)\tau_i + [\dot{x}_i(t + \Delta t)]^2/(2B_i) - [\dot{x}_j(t)]^2/(2B_j) + l_j\}$, where $\dot{x}_i(t + \Delta t) = \dot{x}_i(t) + A_5 \Delta t$, and B_i and B_j are the maximum deceleration rate of i and j , respectively. The astute reader immediately recognizes that the first choice of the right-hand side follows the rationale of the Forbes model [53, 54, 66] and the second choice is similar to that of the Gipps model [57] if driver i is willing to apply an emergency brake (i.e., $b_i = B_i$) as well.

The CARSIM model is compatible with the longitudinal control model. A3 results when \dot{x}_i is set to zero in Equation 21.15. A1 is obtained when the vehicle interaction term (i.e., F_i^j) becomes zero. As vehicle i speeds up, Equation 21.15 predicts that the actual acceleration decreases, which is reflected in lookup Tables 1 and 2 in Refs. [65]. A3 is found when \dot{x}_i is equal to v_i in Equation 21.15. A4 and A5 are related to the longitudinal control model through safety rules which are the same in both models except for a slight implementation difference.

Psychophysical Model

The model developed by Wiedemann [64] is a typical psychophysical model whose principle is depicted in Figure 23.2. The rough curve ABCDEF is a trajectory of the vehicle operation condition in the Δx - $\Delta \dot{x}$ plane. Δx is the spacing between the subject vehicle i and its leader j —that is, $\Delta x = s_{ij}$ —and $\Delta \dot{x}$ is their speed difference $\dot{x}_i - \dot{x}_j$. Starting with point A, vehicle i moves freely if it is not impeded by leader j , which is slower but far ahead. Hence, $\Delta \dot{x}$ remains approximately constant and Δx keeps decreasing. The free-flow state continues up to point B, where the trajectory intersects the

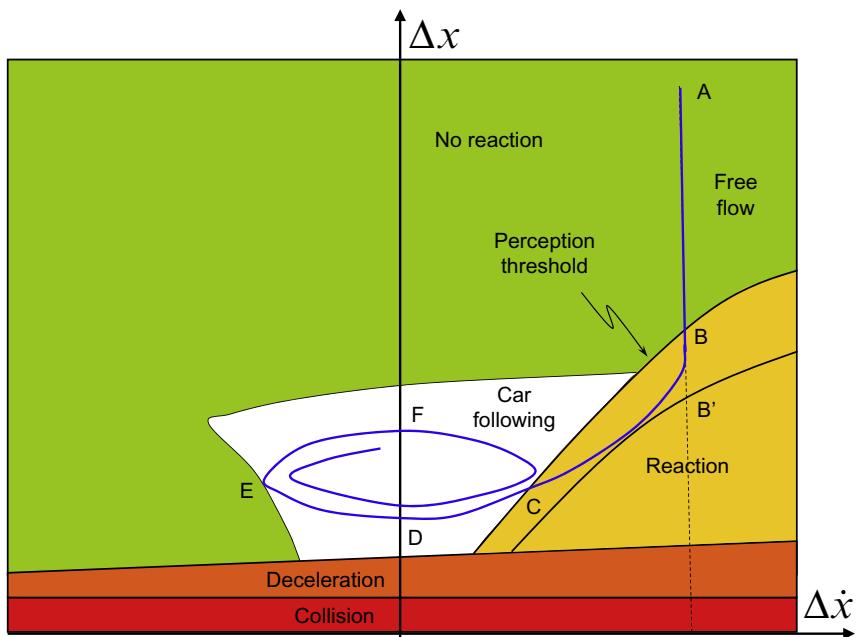


Figure 23.2 The Wiedmann model.

perception threshold. After point B, vehicle i begins to approach vehicle j . In response, driver i reduces his or her speed while Δx keeps decreasing. The approaching regime continues up to point C, where the two vehicles become sufficiently close and their speed difference is small. After point C, the two vehicles are in the car-following regime. As driver i tries to adapt to vehicle j 's speed, the gap closes. Driver i stops decelerating when the two vehicles are moving at the same speed and their distance remains constant. If driver i temporarily loses attention (e.g., talking on a cell phone) and slows down, the gap begins to open until the driver realizes that he or she is falling behind. Consequently, the driver tries to catch up, and hence the gap closes again. If driver i overshoots, he or she may be reminded to back up again. Therefore, the trajectory of driver i oscillates within a unconscious reaction region (the white region) in the Δx - $\Delta \dot{x}$ plane.

Though the psychophysical model is not directly contained in or derived from the field theory as the above-mentioned models are, the effect of the former can be reproduced by the latter. For example, the follower i is in the free-flow regime when the leader j is far ahead. As i moves close to j , the former will ride up onto j 's potential field, and hence perceive a repelling force F_i^j . This signifies the beginning of the approaching regime. As F_i^j increases, \dot{x}_i adapts to \dot{x}_j . Sooner or later, i will find an equilibrium position around the desired spacing s_{ij}^* where the unsatisfied desire for mobility balances the vehicle interaction force. At this point, vehicle i enters the car-following regime. As i 's directional responsiveness (i.e., γ_i) drifts over time, the vehicle may oscillate around the equilibrium position unconsciously, as predicted by [Equation 23.6](#).

Rule-Based Model

The model developed by Kosonen [67] is a representative of rule-based models, and it is reproduced below to be consistent in notation:

1. NO SPEED CHANGE

Keep the present speed (default case).

2. ACCELERATE IF $[\dot{x}_i < v_i]$ and $[t - t_{\text{last}} > T_{\text{acc}}(\dot{x}_i)]$

The current speed \dot{x}_i is less than the desired speed v_i and the time elapsed since the last acceleration t_{last} is more than T_{acc} .

3. NO ACCELERATION IF $[s_{ij} < s_{\min}(\dot{x}_i, \dot{x}_j) + w_{\text{stab}}(\dot{x}_i, \dot{x}_j)]$

The distance from obstacle s_{ij} is less than the minimum safe distance s_{\min} plus the width of the stable area w_{stab} .

4. SLOW DOWN IF $[s_{ij} < s_{\min}(\dot{x}_i, \dot{x}_j)]$

The distance from the obstacle D_{obs} is less than the minimum safe distance s_{\min} .

5. DO NOT SLOW DOWN IF $[\dot{x}_i < \dot{x}_j]$ or $[t - t_{\text{last}} < T_{\text{maxdec}}]$

Own speed is less than obstacle speed or maximum deceleration rate is exceeded.

6. GOTO ZERO IF $[s_{ij} < 0]$ and (Obstacle = physical)

Distance to physical obstacle is below zero (i.e., collision).

At each time step, the motion of vehicle i is checked against the above rules one by one. A later rule always supersedes any earlier rules should there be a conflict. Similarly to the situation in the psychophysical model, the above rule-based model is not directly contained in or derived from the field theory. However, the effect of the rule-based model can be reproduced as well if one is willing to fuzzify the field theory. For example, after fuzzification and discretization, the desired spacing s_{ij}^* can be decomposed into two portions s_{\min} and w_{stab} (see Figure 23.3) to mimic the original setup in Ref. [67]. Therefore, vehicle i does nothing by default if it has reached its desired speed and the road is free (i.e., rule 1 above). If vehicle i 's desire for mobility has not been fully satisfied (i.e., $\dot{x}_i < v_i$), it will accelerate (rule 2). If i approaches j and is within w_{stab} , i will not accelerate (rule 3). Vehicle i needs to decelerate if it intrudes into s_{\min} (rule 4). There is no need for i to decelerate if it becomes slower than j (rule 5). Vehicle i will stop if it collides with j , which is ensured by the steep potential field when the vehicles touch (rule 6).

23.2.3 Relating Macroscopic Equilibrium Models

Under equilibrium condition, vehicles move in a uniform manner and hence lose their identities: $\tau_i \rightarrow \tau$, $\dot{x}_i = \dot{x}_j \rightarrow v$, $v_i = v_j \rightarrow v_f$, $\ddot{x}_i = \ddot{x}_j = 0$, $s_{ij} \rightarrow s = \frac{1}{k}$, $s_{ij}^* \rightarrow s^* = \frac{1}{k^*}$, $b_i = b_j \rightarrow b$, $B_i = B_j \rightarrow B$, and $l_i = l_j \rightarrow \frac{1}{k_j}$, where the right arrow means “aggregate to” and items before the arrow are

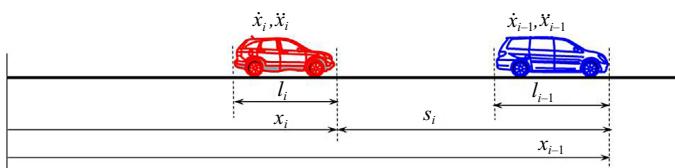


Figure 23.3 The Kosonen model.

microscopic variables and those after the arrow are macroscopic variables. With this notation, the microscopic longitudinal control model (Equation 21.15) translates to its macroscopic counterpart ([Equation 23.11](#) or more specifically [Equations 23.12](#) and [23.13](#)), which depicts the equilibrium speed-density relationship. The field theory and the longitudinal control model are related to existing equilibrium models as follows.

Newell Model (Macroscopic)

If we apply the above notation, the Newell car-following model translates to its macroscopic counterpart of the following form:

$$\nu = \nu_f [1 - e^{\frac{\lambda}{\nu_f} \frac{1}{k_j} (1 - \frac{k_j}{k})}]. \quad (23.21)$$

Notice the close resemblance between [Equations 23.21](#) and [23.11](#). In addition, through its microscopic counterpart, the above model's connection to the longitudinal control model was discussed in [Section 23.2.2](#).

Van Aerde Model (Macroscopic)

The equilibrium counterpart of the Van Aerde model can be written as

$$k = \frac{1}{c_1 + c_3 \nu + c_2 / (\nu_f - \nu)} m \quad (23.22)$$

where all variables are as defined before. Through its microscopic counterpart, the above model is connected to the longitudinal control model as discussed in [Section 23.2.2](#).

IDM (Macroscopic)

Under equilibrium conditions, a special macroscopic case was derived from the IDM [[60](#), [61](#)]:

$$\nu = \frac{(s - L)^2}{2\nu_f T^2} \left[-1 + \sqrt{1 + \frac{4T^2\nu_f^2}{(s - L)^2}} \right], \quad (23.23)$$

where T is average safe time headway and $s = 1/k$ is average spacing, where k is traffic density.

Pipes-Munjal Model

The Pipes-Munjal model [[15](#)] takes the following form:

$$\nu = \nu_f \left[1 - \left(\frac{k}{k_j} \right)^n \right], \quad (23.24)$$

where n is a coefficient and the other variables are as defined before. In the reverse direction (i.e., from macroscopic to microscopic¹), the above model seems to suggest a microscopic basis of roughly the following form:

$$\ddot{x}_i = g_i \left[1 - \frac{\dot{x}_i}{v_i} - \left(\frac{l}{s_{ij}} \right)^n \right]. \quad (23.25)$$

Note that the microscopic basis may take many other forms and the above form is only one of the possibilities. With the above equation, it becomes clear that the Pipes–Munjal model can be derived from the field theory if one chooses the vehicle interaction force F_i^j of the form $(\frac{L}{s_{ij}})^n$. A similar technique can be applied to other equilibrium models in an effort to restore their microscopic basis from the perspective of the field theory.

Drew Model

The Drew model [13] takes the following form:

$$v = v_f \left[1 - \left(\frac{k}{k_j} \right)^{n+\frac{1}{2}} \right], \quad (23.26)$$

where all variables are as defined before. If we repeat the above technique and replace n with $n + \frac{1}{2}$, the suggested microscopic basis is

$$\ddot{x}_i = g_i \left[1 - \frac{\dot{x}_i}{v_i} - \left(\frac{l}{s_{ij}} \right)^{n+\frac{1}{2}} \right], \quad (23.27)$$

which we can derive from the field theory by choosing $F_i^j = g_i (\frac{l}{s_{ij}})^{n+\frac{1}{2}}$.

Model of Wang et al.

Wang et al. [130] recently proposed a stochastic equilibrium model whose three-parameter deterministic version takes the form

$$v = \frac{v_f}{1 + e^{\frac{k-k_c}{\theta}}}, \quad (23.28)$$

where k_c is the critical density (i.e., the density after which speed drop becomes noticeable as density increases from 0 to k_j) and θ is a coefficient. The microscopic basis of the model could be

¹ The same technique was used to derive the Van Aerde car-following model (microscopic) from the Greenshields model (macroscopic).

$$\ddot{x}_i = g_i \left[1 - \frac{\dot{x}_i}{v_i} - \left(1 - \frac{1}{1 + e^{\frac{1}{\theta} \left(\frac{1}{s_{ij}} - \frac{1}{s^c} \right)}} \right) \right] \quad (23.29)$$

where $s^c = 1/k_c$ is the critical spacing (i.e., average spacing at critical density). According to the field theory, one need only choose $F_i^j = g_i [1 - 1/(1 + e^{\frac{1}{\theta}(\frac{1}{s_{ij}} - \frac{1}{s^c})})]$ to obtain the model of Wang et al.

Model of del Castillo and Benítez

Del Castillo and Benítez [118, 131] proposed a family of exponential generating functions which can be represented as

$$f(\lambda) = e^{1-(1+\frac{\lambda}{n})^n}, \quad (23.30)$$

where λ is called the “equivalent spacing,” which is a function of density k , and n is a parameter. Setting $n = 1$ and $n \rightarrow \infty$ results in the following two special cases, respectively:

$$v = v_f \left[1 - e^{\frac{|C_j|}{v_f} (1 - \frac{k_j}{k})} \right] \quad (23.31)$$

and

$$v = v_f \left[1 - e^{1 - e^{\frac{|C_j|}{v_f} \left(\frac{k_j}{k} - 1 \right)}} \right], \quad (23.32)$$

where C_j is the kinematic wave speed at the jam density and the other variables are as defined before. [Equation 23.32](#) is referred to as the “maximum sensitivity curve.” [Equation 23.31](#) takes a form similar to the Newell model and the longitudinal control model. If one chooses $|C_j| = \lambda/k_j$, [Equation 23.31](#) becomes the Newell model, and hence is connected to the longitudinal control model. If the conjecture that $\lambda = 1/\tau$ is true, $|C_j| = \lambda/k_j = l/\tau$, which is the speed required to traverse a nominal vehicle length L (i.e., a vehicle length plus some buffer space) during one perception-reaction time τ . l typically ranges from 5 to 10 m and τ is around 1 s. This yields $|C_j|$ around 5–10 m/s or 11–22 miles per hour, which agrees well with the numbers provided in Ref. [118].

Note that the above two special cases are derived from the exponential family of speed-density curves, which represent a much broader set of models than the Newell model. In addition, the family of speed-density curves can be represented generically as

$$\nu = \nu_f [1 - e^{\psi(k)}], \quad (23.33)$$

where $\psi(k)$ is a generic function and admits the corresponding terms in Equations 23.31 and 23.32. From the perspective of the field theory, the model of del Castillo and Benítez seems to suggest a vehicle interaction force F_i^j proportional to $e^{\psi(1/s_{ij})}$.

GM-associated Models

In addition, the family of equilibrium models, including the models of Greenshields [9], Greenberg [10], Underwood [11], and Drake et al. [12], which are associated with GM models, was discussed in Chapter 14.

23.3 THE UNIFIED DIAGRAM

To summarize the discussion above, a unified perspective can be cast on these traffic flow models. Such a perspective is presented as a diagram in Figure 23.4.

23.3.1 Description of the Unified Diagram

The diagram consists of three panes. The left pane contains *picoscopic* models, which are able to represent vehicle motion in longitudinal x , lateral y , and vertical z directions on a three-dimensional surface. The field theory formulated in Equation 23.6 belongs to this category. The middle pane has *microscopic* car-following models, which describe only vehicle motion in the longitudinal x direction. In this category, models which describe vehicle motion based on acceleration are grouped as “dynamic” models, such as GM models, while those describing vehicle motion based on speed choices are grouped as “steady-state” models, such as the Newell model. The right pane includes *macroscopic* models, which describe equilibrium speed-density relationships. The connecting lines show which models are related. The numbers on these lines, which are explained below, indicate where the bridges between models are discussed in the text. For example, connection 10 indicates the relation between the microscopic longitudinal control model (Equation 21.15) and the Newell nonlinear car-following model (Equation 23.1).

23.3.2 Connections in the Unified Diagram

This subsection refers the connection numbers in Figure 23.4 to the proper locations in the text where the nature of these connections is discussed.

The unified diagram of traffic flow models

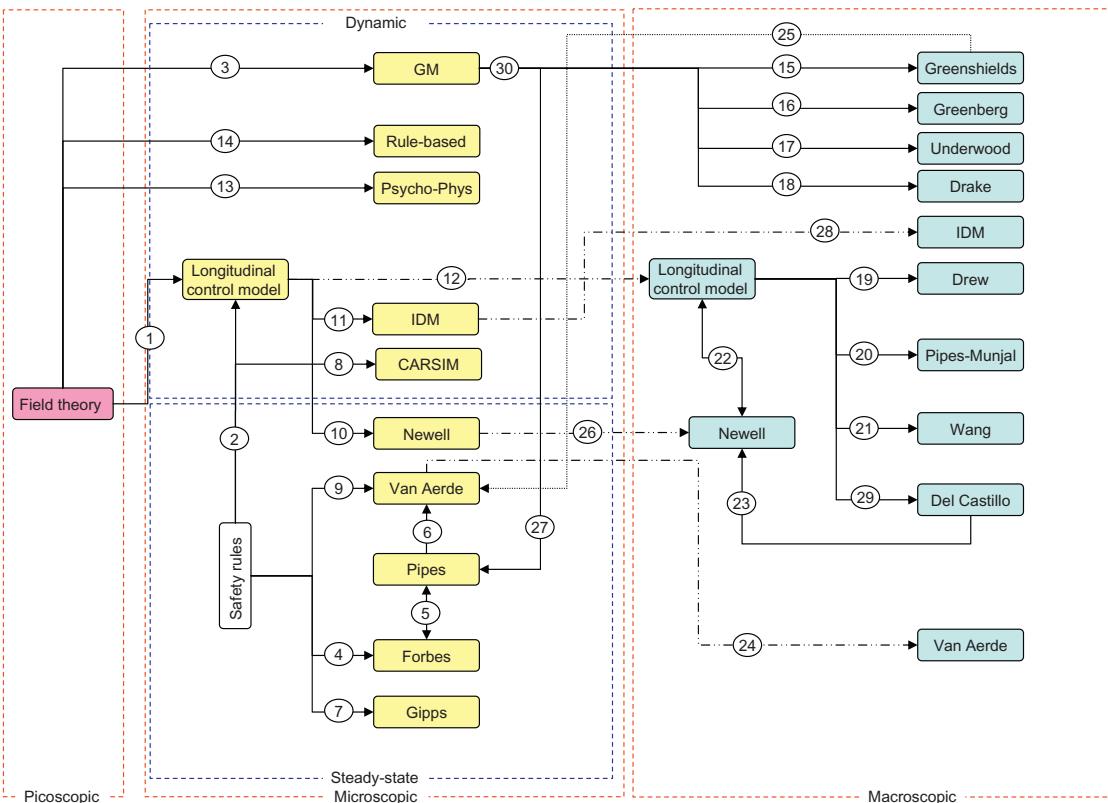


Figure 23.4 The united diagram.

- 1: See [Section 23.2.1](#) for the longitudinal control model as a special case of the field theory.
- 2: See [Section 23.2.1](#) for safety rules being admitted into the longitudinal control model.
- 3: See [Section 23.2.2](#) for GM models as a special case of the field theory.
- 4: See [Sections 23.2.1](#) and [23.2.2](#) for the Forbes model as a safety rule.
- 5: See Chapter 14 for the equivalence between the Pipes model and the Forbes model.
- 6: See [Section 23.2.2](#) for the Pipes model being admitted into the Van Aerde model.
- 7: See [Sections 23.2.2](#) and [§ 23.2.1](#) for the Gipps model as a safety rule.
- 8: See [Section 23.2.2](#) for the relation between the CARSIM model and the longitudinal control model as well as the safety rule in the CARSIM model.
- 9: See [Section 23.2.2](#) for the Van Aerde model as a safety rule.
- 10: See [Sections 23.1](#) and [23.2.2](#) for the Newell model as a special case of the longitudinal control model.
- 11: See [Section 23.2.2](#) for the relation between the IDM and the longitudinal control model.
- 12: See [Section 23.2.1](#) for the derivation of the macroscopic counterpart of the longitudinal control model.
- 13: See [Section 23.2.2](#) for the relation between the psychophysical model and the longitudinal control model.
- 14: See [Section 23.2.2](#) for the relation between the rule-based model and the longitudinal control model.
- 15: See Chapter 14 for the Greenshields model being derived from GM5.
- 16: See Chapter 14 for the Greenberg model being derived from GM5.
- 17: See Chapter 14 for the Underwood model being derived from GM5.
- 18: See Chapter 14 for the model of Drake et al. being derived from GM5.
- 19: See [Section 23.2.3](#) for the relation between the model of Drew et al. and the longitudinal control model.
- 20: See [Section 23.2.3](#) for the relation between the Pipes-Munjal model and the longitudinal control model.
- 21: See [Section 23.2.3](#) for the relation between the model of Wang et al. and the longitudinal control model.
- 22: See [Section 23.2.3](#) for the close resemblance between the Newell model (macroscopic) and the longitudinal control model.

- 23: See [Section 23.2.3](#) for the equivalence between the Newell model (macroscopic) and one of the special cases derived from del Castillo and Benítez's family of exponential generating functions.
- 24: See [[62](#), [63](#)] for microscopic and macroscopic versions of the Van Aerde model.
- 25: See [Section 23.2.2](#) for the Greenshields model being admitted into the Van Aerde model.
- 26: See [Section 23.2.3](#) for the derivation of the macroscopic counterpart of the Newell car-following model.
- 27: See Chapter 14 for the Pipes model being derived from GM5.
- 28: See [Section 23.2.3](#) for the macroscopic IDM being derived from its microscopic counterpart.
- #29: See § [23.2.3](#) for how the Field Theory is related to Del Castillo model.
- #30: See Chapter 14 for how May's original unifying effort fits into the larger Unified Diagram.

23.4 SUMMARY

Motivated by Newell's untold secret in his nonlinear car-following model and May's original unifying effort depicted in Figure 6.6 in Ref. [[17](#)], a broader unified perspective was cast on traffic flow modeling and a larger unified diagram was constructed.

The Newell model [[58](#)], after being rearranged slightly, gives rise to a mechanics model which involves noncontact forces, which, in turn, can be explained conveniently using the concept of a field. The field theory of this nature was presented in Chapter 21, and was concisely reproduced here for easy reference. With use of the field theory as a framework, existing traffic flow models can be related to each other, thereby providing a unified perspective to examine the coherence among these models.

Microscopic car-following models are related to the field theory by variation of its components, such as vehicle interaction force, desired spacing (via safety rules), and directional responsiveness. Even though some models are not directly contained in or derived from the field theory, their effects can be reproduced from the latter. When aggregated, many of these car-following models reduce to their macroscopic counterparts—that is, equilibrium speed-density relationships. Those macroscopic equilibrium models, which do not come with a proposed microscopic basis, fortunately

contain information to deduce their microscopic nature, though which the connection to the field theory might be established.

To summarize the above analysis, a unified diagram was constructed which gathers together traffic flow models at the picoscopic, microscopic, and macroscopic levels with lines connecting related models. Each connection is denoted by a number which points to the discussion of the connection in this book.

PROBLEMS

1. Microscopic car-following models come in different flavors. One type of car-following model uses acceleration as the driver's control variable, and this sounds like instructing the driver to "speed up or slow down by $x \text{ m/s}^2$ next." We call them *dynamic* models. Another type of car-following model employs speed as the control variable, which suggests that the driver ought to "bring the speed to $y \text{ km/h}$ next." We call them *steady-state* models. A third type of car-following model works on the position of the vehicle, which translates to asking the driver to "move your vehicle to position $z \text{ m}$ next." We call them *static* models. Search the car-following models and provide at least one example for each type.
2. When proposing his nonlinear car-following model, Newell acknowledged that "no motivation for this choice is proposed other than the claim that it has approximately the correct shape and is reasonably simple." Give your opinion on this modeling philosophy and discuss whether or not it is a good one to follow.
3. Many macroscopic equilibrium models have been proposed, and their flexibility to fit field observations varies depending partially on the number of parameters they employ. Provide at least three example equilibrium models that employ
 - a. two parameters,
 - b. three parameters, and
 - c. four parameters.

CHAPTER 24

Multiscale Traffic Flow Modeling

Thus far, this book has presented traffic flow theory progressively from macroscopic to microscopic to picoscopic and from the “obvious” field observations to simple equilibrium models to involved dynamic models to complicated driver-vehicle-environment closed-loop systems. It is natural to extend the line of thinking to multiscale modeling, where high-level models provide system-wide overview, while low-level models describe local operation details. In addition, it is critical to adopt a consistent modeling approach to ensure the coupling between different levels.¹

24.1 INTRODUCTION

Anyone who used maps has probably had the following experience. Fifteen years ago, a 1:10,000 paper map was needed to view a city (e.g., Amherst, MA, USA), while a 1:1,000,000 paper map was needed to view a state (e.g., Massachusetts). If the scale was changed, a new map was needed. Today, using digital maps (e.g., Google Maps), one is able to view the entire country, and then progressively zoom in to view Massachusetts, Amherst, and even the University of Massachusetts Amherst campus, all seamlessly and within a single system.

Similarly, it is desirable that traffic simulation allows an analyst to zoom in to examine low-level details and zoom out to view system-wide performance within the same simulation process. Figure 24.1 illustrates such a paradigm. The background represents a *macroscopic* view of traffic operation in an entire region. This is analogous to viewing traffic from 10,000 m above the ground and the traffic appears to be a compressible fluid whose states (speed, flow, density, etc.) propagate like waves. As one zooms in to a local area of the region, a *mesoscopic* view is obtained. This is like viewing traffic from 3000 m above the ground, where the sense of waves recedes and a scene of particles emerges. As one further zooms in to a segment of the roadway, a *microscopic* view results. Similarly to watching traffic from 1000 m above the ground, the scene is dominated by moving particles that

¹ This chapter is reproduced from [114].

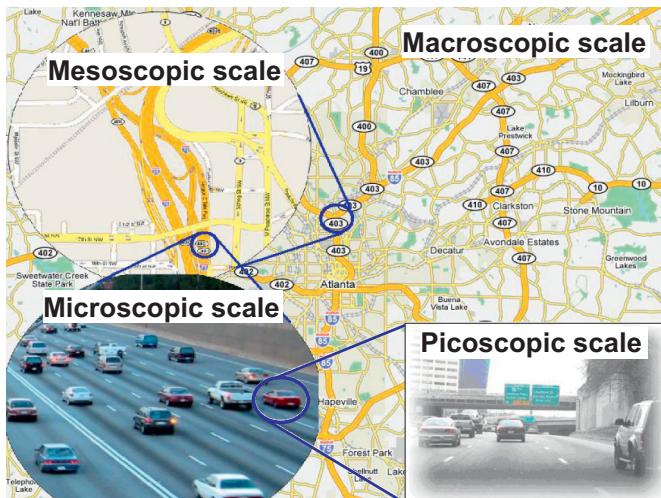


Figure 24.1 Multiscale traffic flow modeling.

interact with each other so as to maintain safe positions in a traffic stream. Finally, if one focuses on a few neighboring vehicles, a *picoscopic* view is achieved as if one were operating one of the vehicles. As such, one has to interact with the driving environment (e.g., roadway, signs, signals), make control decisions, and manage vehicle dynamic responses to travel safely. If such a “zoomable” simulation becomes available, one would be able to translate a traffic flow representation at multiple scales—for example, to trace a low-level event all the way to a high-level representation and, conversely, to decompose a global problem into one or more local deficiencies. As such, the “zoomable” simulation will transform the way that traffic flow is analyzed and transportation problems are addressed.

The objective of this chapter is to address multiscale traffic flow modeling with inherent consistency. The term “consistency” here concerns the coupling among models at different scales—that is, how less detailed models are derived from more detailed models and, conversely, how more detailed models are aggregated to less detailed models. Only consistent multiscale models are able to provide the theoretical foundation for the above “zoomable” traffic simulation. The chapter is organized as follows. [Section 24.2](#) takes a broad perspective on a spectrum of four modeling scales. Modeling objectives and model properties at each scale are discussed, and existing efforts are reviewed. [Section 24.3](#) presents the multiscale approach based on the field theory. The modeling strategy at each scale is discussed,

and some special cases are formulated at both the microscopic scale and the macroscopic scale. The emphasis of this multiscale approach is to ensure coupling among different modeling scales. Concluding remarks and future directions are presented in [Section 24.4](#).

24.2 THE SPECTRUM OF MODELING SCALES

The modeling of traffic flow can be performed at, but is not limited to, four scales—namely, picoscopic, microscopic, mesoscopic, and macroscopic, from the most to the least detailed. Considering that the definitions of these modeling scales are rather vague, implicit, or absent in the literature, this section attempts to provide an explicit definition so that existing and future models are easily classified and related. Such a definition is tabulated in [Figure 24.2](#) for each of the four modeling scales on the basis their properties (i.e., rows in the table), and literature related to each modeling scale is reviewed in subsequent subsections. The first three rows (“state variable,” “variable description,” and “state diagram”) are discussed in this section, and

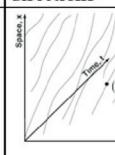
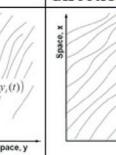
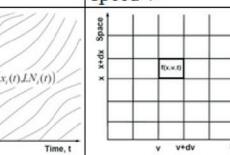
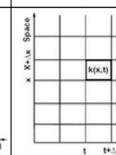
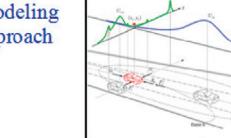
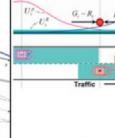
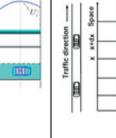
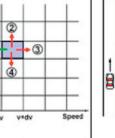
Scale	Picoscopic	Microscopic	Mesoscopic	Macroscopic
State variable	$(x_i(t), y_i(t))$ $i = 1, 2, 3, \dots$ $0 < t < \infty$	$(x_i(t), LN_i(t))$ $LN \in \{1, 2, \dots, n\}$	$f(x, v, t)$	$k(x, t)$
Variable description	Vehicle trajectory in longitudinal x and lateral y directions	Vehicle trajectory in x direction and lane # LN in y direction	Distribution of a vehicle at location x and time t with speed v	Concentration of vehicles at location x and time t
State diagram				
Underlying principle	Control theory System dynamics Field theory	Field theory	Statistical mechanics	Fluid dynamics
Modeling approach				
Model coupling		Pico-Micro	Micro-Meso	Meso- Macro
			Micro - Macro	

Figure 24.2 The spectrum of modeling scales.

the remaining three rows (“underlying principle,” “modeling approach,” and “model coupling”) will be described in the next section.

24.2.1 The Picoscopic Scale

Picoscopic modeling should be able to represent traffic flow so that the trajectory of each vehicle, $(x_i(t), y_i(t))$, where $i \in \{1, 2, 3, \dots, I\}$ denotes the vehicle ID, can be tracked in both the longitudinal x direction and the lateral y direction over time $t \geq 0$. Knowing these vehicle trajectories, one can completely determine the state and dynamics of the traffic system. Therefore, $(x_i(t), y_i(t))$ is the state variable (one or a set of variables that characterizes the state of a system). The corresponding state diagram (a graphical representation that illustrates the dynamics or evolution of system state) consists of these vehicle trajectories in a three-dimensional domain (x, y, t) .

Picoscopic models are mainly of interest in automotive engineering. Dynamic vehicle models with varying degrees of freedom have been proposed [132, 133]. A myriad of driver models have been reported to assist various aspects of automotive engineering, including vehicle handling and stability. Control theory was widely applied in modeling vehicle control [134, 135]. Models in this category typically incorporate one or more feedback loops. These loops are used by the controller to adjust its output to minimize control error. Human drivers can better perform reasoning using vague terms than can controllers. This observation allows the use of fuzzy logic [136, 137], which controls vehicles on the basis of some predefined rules. To allow implicit driving rules, artificial neural networks [138, 139] learn “driving experiences” from training processes and then apply the learned experiences in future driving. Several literature surveys of driver models are available [140–142].

24.2.2 The Microscopic Scale

Microscopic modeling should be able to represent traffic flow so that the trajectory of each vehicle can be tracked in the longitudinal direction $x_i(t)$, with the lateral direction being discretized by lanes $\text{LN}_i(t)$ where $\text{LN} \in \{1, 2, \dots, n\}$. Hence, $(x_i(t), \text{LN}_i(t))$ is a state variable that describes the state and dynamics of traffic flow at this scale, and the corresponding state diagram consists of vehicle trajectories in a two-dimensional domain (x, t) .

Within the traffic flow community, microscopic models treat driver-vehicle units as massless particles with personalities. The behavior of these particles is governed by car-following models in the longitudinal direction and discrete-choice (e.g., lane-changing and gap-acceptance) models in

the lateral direction. Car-following models describe how a vehicle (the follower) responds to the vehicle in front of it (the leader). For example, stimulus-response models [55, 56] assume that the follower's response (e.g., desired acceleration) is the result of stimuli (e.g., spacing and relative speed) from the leader, desired measure models [52, 57] assume that the follower always attempts to achieve his desired gains (e.g., speed and safety), psychophysical models [64, 111] introduce perception thresholds that trigger driver reactions, and rule-based models [67] apply “IF-THEN” rules to mimic driver decision making. Lane-changing and gap-acceptance models describe how a driver arrives at a lane change decision and how the driver executes such a decision, respectively. Approaches to lane changing include mandatory and discretionary lane changing [143, 144], adaptive acceleration mandatory and discretionary lane changing [145, 146], and autonomous vehicle control [147]. The following have been attempted to model gap acceptance: deterministic models [148–150], probabilistic models [151–153], and neuro-fuzzy hybrid models [154]. More surveys on microscopic models can be found in the literature [3, 155].

24.2.3 The Mesoscopic Scale

Mesoscopic modeling should be able to represent traffic flow so that the probability of the presence of a vehicle at a longitudinal location x with speed v at time t is tracked. The lateral direction is of interest only if it provides passing opportunities. The state diagram typically involves a two-dimensional domain (x, v) at an instant t , and the domain is partitioned into cells with space increment dx and speed increment dv . The state variable is a distribution function $f(x, v, t)$ such that $f(x, v, t)dx dv$ denotes the probability of having a vehicle within space range $(x, x + dx)$ and speed range $(v, v + dv)$ at time t . Knowing the distribution function $f(x, v, t)$, one can determine the dynamics of the system statistically.

Conventional mesoscopic traffic flow models come in three flavors. First, models such as the one in TRANSIMS [156] take a cellular automata approach, where the space domain (representing the longitudinal direction of a highway) is partitioned into short segments typically 7.5 m long. If it is occupied, a segment is able to store only one vehicle. Vehicles are then modeled as hopping from one segment to another, so their movement and speed are discretized and can take only some predetermined values. Second, models such as those implemented in DynaMIT [157] and DYNASMART [158] keep track of the motion of individual vehicles, but their speeds are determined with use of macroscopic models (such as an equilibrium

speed-density relationship) instead of microscopic car-following models. Third, truly mesoscopic models such as the one postulated by Prigogine and his coworkers [159] are based on nonequilibrium statistical mechanics or kinetic theory, which draws an analogy between classical particles and highway vehicles. However, Prigogine's model was criticized [160] for (1) lacking a theoretical basis, (2) lacking realism (e.g., car following, driver preferences, and vehicle lengths), and (3) lacking satisfactory agreement with empirical data. Many efforts have been made to improve Prigogine's model by addressing criticisms 2 and 3. For example, Paveri-Fontana [161] considered a driver's desired speeds, Helbing [162] adapted the desired speeds to speed limits and road conditions, Phillips [39, 163] incorporated vehicle lengths, Nelson [164] accounted for vehicle acceleration behavior, and Klar and Wegener [165, 166] included a stochastic microscopic model. Surveys of existing approaches are available in Ref. [167].

24.2.4 The Macroscopic Scale

Macroscopic modeling should be able to represent traffic flow so that only local aggregation of traffic flow (e.g., density k , speed u , and flow q) over space (longitudinal) x and time t is tracked. Traffic density $k(x, t)$ is a good candidate of state variable because, unlike flow and speed, density is an unambiguous indicator of the traffic condition. The state diagram typically involves a two-dimensional domain (x, t) . Knowing $k(x, t)$, one can determine the dynamics of the system macroscopically.

Conventional macroscopic traffic flow models describe the propagation of traffic disturbances as waves. A fundamental basis for formulating wave propagation is the law of conservation. The first-order form of the law is mass/vehicle conservation, which is used to create first-order models [24, 25]. In addition, momentum and energy are other forms of conservation. A model is of a higher order if it incorporates the latter forms of conservation [37, 38]. Since the limited benefit offered by higher-order models often does not justify their added complexity [47], numerical approximation and macroscopic simulation have been centered on first-order models—for example, KRONOS [27], the kinematic waves model [31], the cell transmission model [28, 29], FREQ [26], and CORQ [168]. More surveys of macroscopic models can be found in the literature [3].

24.2.5 Issues of Multiscale Modeling

Remarkably, existing models at the same scale typically follow different modeling approaches, and hence it is difficult to relate these models to each

other. In addition, models at different modeling scales are rarely coupled. For example, a macroscopic model typically lacks a microscopic basis, and a microscopic model does not have its macroscopic counterpart.

Therefore, an ideal multiscale modeling approach should emphasize not only model quality at each individual scale but also the coupling between different scales. Only models formulated with such an approach are able to support the “zoomzble” traffic simulation discussed in [Section 24.1](#). As such, the resulting state diagram at a more detailed scale contains the necessary information to reproduce a less detailed diagram, as illustrated in [Figure 24.2](#). For example, the microscopic diagram is simply a projection of the picoscopic diagram onto the x - t plane, and the macroscopic state diagram can be completely reconstructed from the microscopic diagram with use of Edie’s definition of traffic flow characteristics [4, 6].

24.3 THE MULTISCALE APPROACH

The objective of this section is to pursue the above multiscale modeling approach and develop strategies to formulate a spectrum of models with inherent consistency. The approach starts at the picoscopic scale by formulating a model that is mathematically amenable to representing the natural way of human thinking while complying with physical principles; the microscopic model can be simplified from the picoscopic model yet still captures the essential mechanisms of vehicle motion and interaction; the mesoscopic model can be derived from the microscopic model on the basis of principles of nonequilibrium statistical mechanics; the macroscopic model can be derived from the mesoscopic model by application of the principles of fluid dynamics. See [Figure 24.2](#) for a summary of the underlying principles, the modeling approaches, and modeling coupling.

24.3.1 Picoscopic Modeling

This section consolidates and highlights the presentation in Chapters 18 and 21 as follows. To conform to real-world driving experiences, the picoscopic model should mimic the way that a driver operates his/her vehicle and responds to the driving environment. On the basis of the principles of control theory, a driver-vehicle-environment closed-loop control system has been developed. [Figure 24.3](#) illustrates the components of the system and its control flow, including feedback loops.

This system consists of a driver model and a vehicle model which interact with each other as well as with the driving environment. The

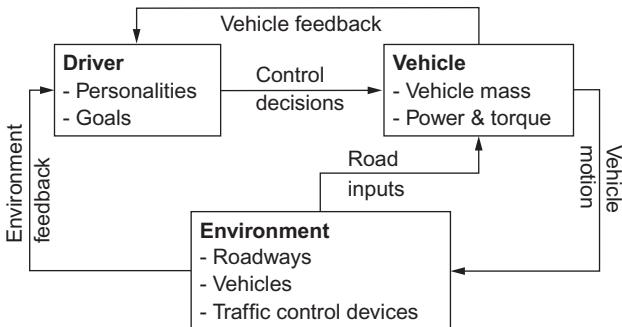


Figure 24.3 The closed-loop system.

driver receives information from the environment, such as roadways, traffic control devices, and the presence of other vehicles. The driver also receives information from his/her own vehicle, such as speed, acceleration, and yaw rate. These sources of information, together with driver properties and goals, are used to determine driving strategies (such as steering and accelerating/braking). The driving strategies are fed forward to the vehicle, which also receives input from roadways. These sources of information, together with vehicle properties, determine the vehicle's dynamic responses based on vehicle dynamic equations. Moving longitudinally and laterally, the vehicle constitutes part of the environment. Other vehicle dynamic responses such as speed, acceleration, and yaw rate are fed back to the driver to determine driving strategies in the next step. Thus, traffic operation is composed of movement and interaction of all vehicles in the environment.

The driver model can be formulated by applying the principles of the field theory. Basically, objects in a traffic system (e.g., roadways, vehicles, and traffic control devices) are perceived by a driver as component fields. The driver interacts with an object at a distance, and the interaction is mediated by the field associated with the object. The superposition of these component fields represents the overall hazard encountered by the driver. Hence, the driving strategy is to seek the least hazardous route by navigating through the field along its valley, and traffic flow consists of the motion and interaction of all vehicles. With this understanding, the driver model at the picoscopic scale is formulated as follows.

The driver's strategy of moving on roadways is to achieve gains (mobility and safety) and avoid losses (collisions and violation of traffic rules). Such a strategy can be represented as navigating through the valley of an overall

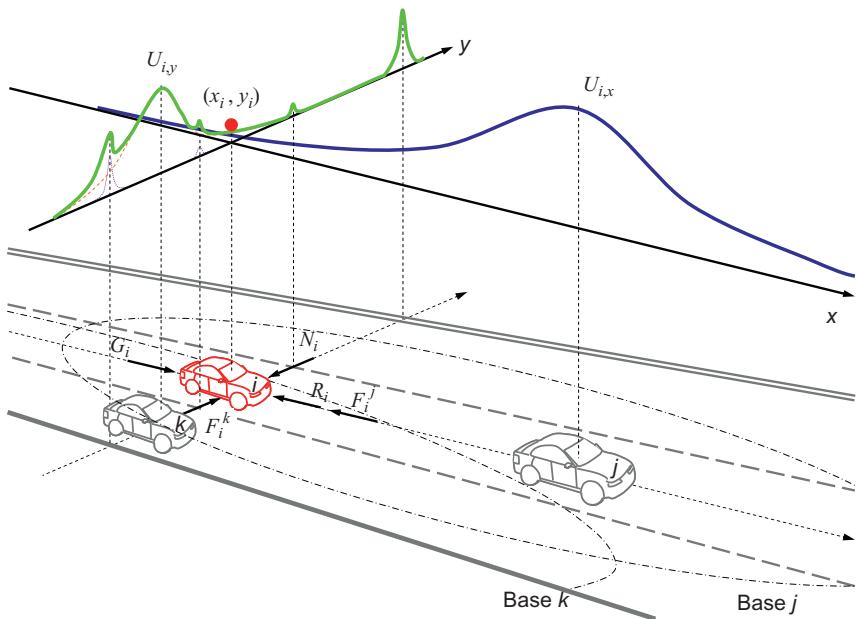


Figure 24.4 The illustration of a perceived field.

field U_i which consists of component fields such as those due to moving units U_i^B , roadways U_i^R , and traffic control devices U_i^C —that is,

$$U_i = U_i^B + U_i^R + U_i^C.$$

For example, Figure 24.4 illustrates two sections of the overall field, $U_{i,x}$ and $U_{i,y}$. The subject vehicle i is represented as a ball which rides on the tail of curve $U_{i,x}$ since the vehicle is within vehicle j 's field. Therefore, vehicle i is subject to a repelling force F_i^j which is derived from $U_{i,x}$ as

$$F_i^j = -\frac{\partial U_{i,x}}{\partial x}.$$

The effect of F_i^j is to push vehicle i back to keep a safe distance. By incorporating the driver's unsatisfied desire for mobility ($G_i - R_i$), we can determine the net force in the x direction as

$$m_i \ddot{x}_i = \sum F_{i,x} = G_i - R_i - F_i^j = (G_i - R_i) + \frac{\partial U_{i,x}}{\partial x}.$$

The section of U in the lateral γ direction, $U_{i,y}$ (the bold curve), is the sum of two components: the cross section of the field due to vehicle k (the dashed curve) and that due to the roadway field (the dotted curve). The former results in a repelling force F_i^k which makes driver i shy away from k and the latter generates a correction force N_i if i deviates from its lane center. Therefore, the net effect can be expressed as:

$$m_i \ddot{y}_i = \sum F_{i,y} = F_i^k - N_i = -\frac{\partial U_{i,y}}{\partial \gamma}.$$

By incorporating time t , driver i 's perception-reaction time τ_i , and driver i 's directional response γ , we can express the above equations as

$$\begin{aligned} m_i \ddot{x}_i(t + \tau_i) &= \sum \tilde{F}_{i,x}(t) = \gamma_i^0 [G_i(t) - R_i(t)] + \gamma(\alpha_i^j) \frac{\partial U_{i,x}}{\partial x}, \\ m_i \ddot{y}_i(t + \tau_i) &= \sum \tilde{F}_{i,y}(t) = -\gamma(\alpha_i^k) \frac{\partial U_{i,y}}{\partial \gamma}, \end{aligned}$$

where $\gamma_i^0 \in [0, 1]$ represents the driver's attention to unsatisfied desire for mobility (typically $\gamma_i^0 = 1$), and α_i^j , α_i^k , and α_i^N are viewing angles, which are also functions of time.

24.3.2 Microscopic Modeling

We can formulate the microscopic model by simplifying the above pico-scopic model as follows: (a) ignoring interactions inside a driver-vehicle unit, allowing it to be modeled as an active particle, (b) representing a driver's longitudinal and lateral control using separate but simpler models, (c) reducing the vehicle dynamic system to a particle, and (d) simplifying road surface and lanes to a collection of parallel lines.

Modeling Longitudinal Control

With the above simplifications, the three-dimensional potential field U in [Figure 24.4](#) reduces to a two-dimensional potential function. The upper part of [Figure 24.5](#) illustrates an example where driver i (the middle one) is traveling behind a leading vehicle j and is followed by a third vehicle p in the adjacent lane. The potential field U_i perceived by the driver is shaded in the lower part [Figure 24.5](#) and is represented by a curve in the upper part. Since the trailing vehicle in the adjacent lane does not affect the subject driver's longitudinal motion, the “stress” on the subject driver to keep a safe distance comes only from the leading vehicle and can be represented as

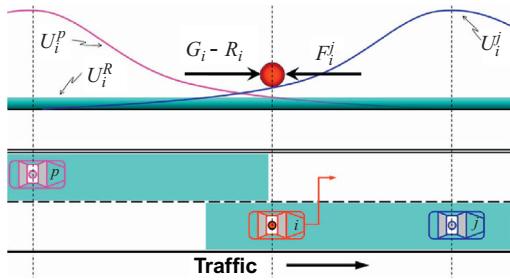


Figure 24.5 Microscopic modeling.

$$F_i^j = -\frac{\partial U_i^j}{\partial x}.$$

By incorporating roadway gravity G_i , roadway resistance R_i , and interaction between vehicles F_i^j , we can express the net force on i more specifically as

$$m_i \ddot{x}_i = G_i - R_i - F_i^j.$$

If one chooses proper functional forms for the above terms, special cases of the model can be obtained—for example, the longitudinal control model presented in Chapter 22:

$$\ddot{x}_i(t + \tau_i) = g_i \left[1 - \left(\frac{\dot{x}_i(t)}{v_i} \right) - e^{-\frac{s_{ij}(t)^* - s_{ij}(t)}{s_{ij}(t)^*}} \right], \quad (24.1)$$

$$s_{ij}^*(t) = x_{i-1}(t) - x_i(t) \geq \frac{\dot{x}_i^2(t)}{2b_i} + \dot{x}_i \tau_i - \frac{\dot{x}_{i-1}^2(t)}{2B_{i-1}} + l_j,$$

where it is assumed that $G_i = m_i \times g_i$, $R_i = m_i \times (\frac{\dot{x}_i(t)}{v_i})$, and $F_i^j = m_i \times f(s_{ij}, s_{ij}(t)^*)$, where g_i is the maximum acceleration that driver i is willing to apply when starting from standstill, $\dot{x}_i(t)$ is the actual speed of vehicle i , v_i is the desired speed of driver i , $s_{ij} = x_j - x_i$ is the actual spacing between vehicles i and j , x_i is the position of vehicle i , x_j is the position of vehicle j , and s_{ij}^* is the desired spacing between vehicles i and j . l_j is the nominal length of vehicle j and is conveniently used as the spacing between two vehicles in jammed traffic. The difference $(s_{ij}^* - s_{ij})$ represents how far vehicle i intrudes beyond s_{ij}^* . The rationale for representing the interaction force F_i^j between vehicles i and j with an exponential function is to set the desired

spacing s_{ij}^* as a baseline, beyond which the intrusion by vehicle i is translated exponentially to the repelling force acting on the vehicle.

The desired spacing s_{ij}^* is derived according to the Gipps model [57]. More specifically, s_{ij}^* should allow vehicle i to stop behind its leading vehicle j after a perception-reaction time τ_i and a deceleration process at a comfortable level $b_i > 0$ should vehicle j apply an emergency brake at rate $B_j > 0$. Of course, the desired spacing can be derived on the basis of other safety rules if appropriate.

Modeling Lateral Control

The driver's lateral control concerns changing lanes to seek a speed gain or to use an exit. The shaded areas in the bottom part of Figure 24.5 can be interpreted as driver j and p 's personal spaces after the lane barrier has been accounted for. A lane change decision is reached whenever driver i intrudes into another driver's personal space. With such a decision, driver i begins to search for open spaces in adjacent lanes. In this particular case, an open space happens to be available in the left lane, barely allowing the center of vehicle i to move in. Consequently, the result of the gap-acceptance decision is to abruptly switch vehicle i to the left lane.

24.3.3 Mesoscopic Modeling

Mesoscopic modeling applies the principles of nonequilibrium statistical mechanics or kinetic theory to model traffic flow. Essential to the modeling is the determination of a distribution function $f(x, v, t)$ such that $f(x, v, t)dx dv$ denotes the probability of having a vehicle within space range $(x, x + dx)$ and speed range $(v, v + dv)$ at time t (see Figure 24.6). The time evolution of traffic flow is described by an evolution equation,

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x} \frac{dx}{dt},$$

whose right-hand side is to be determined. Therefore, the central question is how to rigorously derive the evolution equation. This can be done by use of a procedure similar to that used to derive the Boltzmann equation [169, 170] from basic principles. The classical Boltzmann equation describes particles moving in a three-dimensional domain, so the first step is to reduce the three-dimensional case to a one-dimensional case which represents traffic moving on a unidirectional highway.

Existing models, in particular those based on Prigogine's work, are postulated. To derive the one-dimensional Boltzmann equation from basic

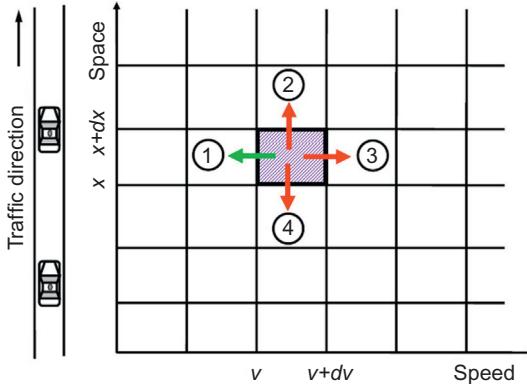


Figure 24.6 The x - v diagram.

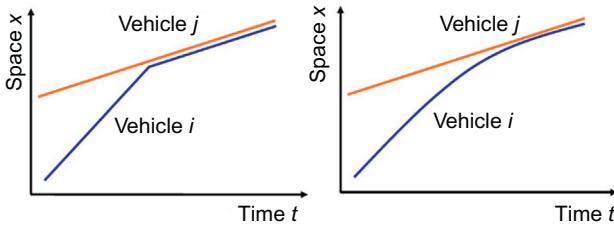


Figure 24.7 Car following.

principles, a sound understanding of the mechanism of traffic evolution is required. Existing models, including a derived model [165, 166], assumed that the mechanism is vehicle “collision.” For example, the fast follower i in the left panel in Figure 24.7 keeps its speed up to the collision point and then abruptly changes its speed. To be realistic, the speed change of vehicle i needs to be smooth as it approaches its leader j as illustrated in the right panel in Figure 24.7. This is possible only if car following is incorporated as the mechanism of particle interaction. As such, the longitudinal control model can be used to derive the one-dimensional Boltzmann equation and, thus, ensures micro-meso coupling.

The derivation of the one-dimensional Boltzmann equation starts from the application of the conservation law (e.g., vehicles entering and exiting the highlighted cell in Figure 24.6 should be conserved). Existing models considered only one direction (i.e., direction 1 below), in which vehicles exit the cell, and a similar treatment applies to vehicles entering the cell.

This approach causes modeling errors. Actually, vehicles may exit the cell in four directions: (1) vehicles slowed down (and hence exited the cell) because of a sluggish leader, (2) vehicles physically moved out of the cell, (3) vehicles accelerated because of an aggressive follower, and (4) vehicles reversed, which is unlikely. The opposite applies to vehicles entering the cell. Therefore, application of the law to include all directions is the correct approach. Since it is mathematically complicated to derive the one-dimensional Boltzmann equation, this chapter presents only potential directions of exploration, leaving the actual derivation to be addressed in future research.

Once the one-dimensional Boltzmann equation has been formulated, one may solve it using initial and boundary conditions to study how traffic evolves over time and space. However, solving the equation can be quite involved, as is the case for any classical Boltzmann equation. Fortunately, some important results can be inferred without fully solving the equation. For example, a hydrodynamical formulation, which is essential to macroscopic modeling, can be derived from the equation. In addition, the equation contains an equilibrium relationship between vehicle speed and traffic density which is also essential to macroscopic modeling. Such a relationship is analogous to the Maxwell-Boltzmann distribution (the distribution of molecular speed at different temperatures) which is the stationary (i.e., $\frac{\partial f}{\partial t} = 0$) solution to a classical Boltzmann equation.

24.3.4 Macroscopic Modeling

Macroscopic modeling applies the principles of fluid dynamics to model traffic flow as a one-dimensional compressible continuum fluid. While the above mesoscopic modeling describes the distribution of vehicles in a highway segment, macroscopic modeling represents only the average state. Therefore, traffic density $k(x, t)$ can be related to the distribution $f(x, v, t)$ as its zeroth moment $k(x, t) = \int f(x, v, t) dv$ and traffic speed as the first moment $u(x, t) = \frac{1}{k} \int vf(x, v, t) dv$. From this understanding, it becomes clear that it is feasible to derive a hydrodynamical formulation from the mesoscopic model. The one-dimensional Boltzmann equation discussed above can be expressed in a general form as

$$\frac{\partial f}{\partial t} + v \frac{\partial f}{\partial x} = C,$$

where C denotes the rate of change of $f(x, v, t)$. Multiplying both sides of this equation by 1, v , and $\frac{1}{2}v^2$ and integrating it over v , we obtain hydrodynamical equations of mass, momentum, and energy conservation. The mass conservation equation

$$\frac{\partial k}{\partial t} + \frac{\partial(ku)}{\partial x} = \int C dv$$

is of particular interest because it describes the time evolution of traffic density $k(x, t)$. To solve the equation, a speed-density relationship must be introduced into the macroscopic model. This relationship can be derived from the mesoscopic model under stationary conditions or, alternatively, can be obtained directly from the microscopic model if equilibrium conditions are assumed. For example, the macroscopic version of the longitudinal control model is

$$v = v_f [1 - e^{1 - \frac{k^*}{k}}], \quad (24.2)$$

where $k^* = \frac{1}{\gamma v^2 + \tau v + l}$, v_f is the free-flow speed, $k_j = 1/l$, l is the bumper-to-bumper distance between vehicles when traffic is jammed, and τ is the average perception-reaction time of drivers.

Therefore, the macroscopic model consists of a system of equations including the hydrodynamical formulation and one of the above speed-density relationships:

$$\frac{\partial k}{\partial t} + \frac{\partial(ku)}{\partial x} = \int C dv,$$

$$v = V(k).$$

We can solve the system of equations graphically using the method of characteristics or numerically using a finite difference approach. A typical finite difference method is illustrated in [Figure 24.8](#), where one partitions the time-space domain into cells and keeps track of traffic flowing into and out of each cell [21, 27, 171].

24.4 SUMMARY

This chapter has presented a broad perspective on traffic flow modeling at four scales: picoscopic, microscopic, mesoscopic, and macroscopic, from the most to the least detailed level. Modeling objectives and model properties at each scale were discussed and existing efforts were reviewed.

To ensure modeling consistency and provide a microscopic basis for macroscopic models, it is critical to address the coupling among models

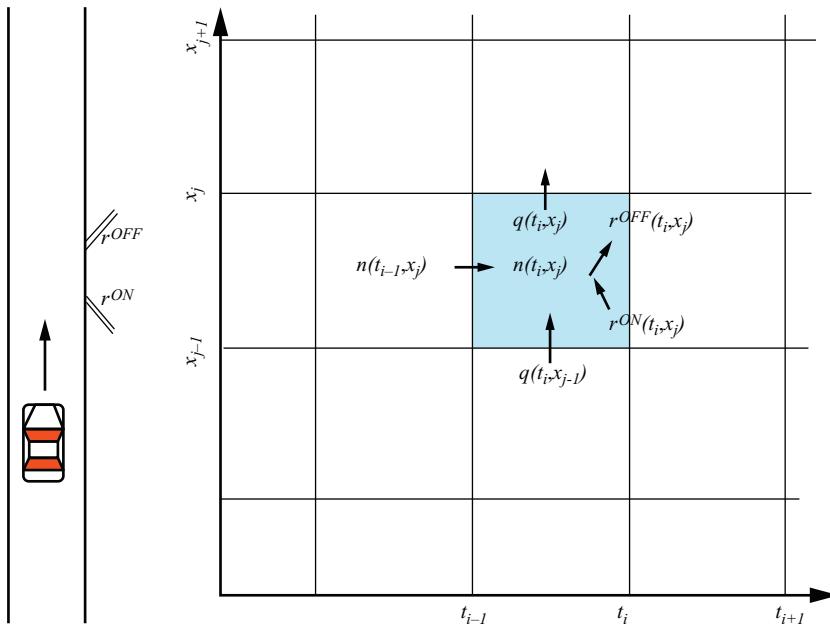


Figure 24.8 The finite difference method.

at different scales—that is, how less detailed models are derived from more detailed models and, conversely, how more detailed models are aggregated to less detailed models. With this understanding, a consistent modeling approach was proposed based on the field theory. Basically, in this approach, objects (e.g., roadways, vehicles, and traffic control devices) are perceived by the subject driver as component fields. The driver interacts with an object at a distance, and the interaction is mediated by the field associated with the object. In addition, the field may vary when perceived by different drivers depending on their characteristics, such as responsiveness and perception-reaction time. The superposition of these component fields represents the overall hazard encountered by the subject driver. Hence, the objective of the driver is to seek the least hazardous route by navigating through the field along its valley. Consequently, traffic flow is modeled as the motion and interaction of all vehicles.

Modeling strategies at each of the four scales were discussed. More specifically, the field theory serves as the basis of picoscopic modeling, which represents a driver-vehicle unit as a driver-vehicle-environment closed-loop control system. The system is able to capture vehicle motion in longitu-

dinal and lateral directions. The microscopic model is obtained from the picoscopic model by simplification of its driver–vehicle interactions, vehicle dynamics, and vehicle lateral motion. The mesoscopic model is derived from basic principles with use of the microscopic model as the mechanism of traffic evolution. The macroscopic model includes an evolution equation (which is derived by taking moments of the mesoscopic model) and an equilibrium speed–density relationship (which is the stationary solution to the mesoscopic model or is derived from the microscopic model directly). Therefore, the proposed approach ensures model coupling and modeling consistency. As such, consistent models derived from this approach are able to provide the theoretical foundation to develop the “zoomable” traffic simulation tool discussed in [Section 24.1](#).

PROBLEMS

1. This chapter discussed a spectrum of four modeling scales—namely, macroscopic, mesoscopic, microscopic, and picoscopic, from the least to the most detailed. Provide at least one example model at each scale.
2. Define each of the four modeling scales with their properties, such as state variables.
3. Discuss the issues that multiscale traffic flow modeling is currently facing in relation to existing models.

BIBLIOGRAPHY

- [1] K. Petty, Small time scale analysis of loop data, <http://ipa.eecs.berkeley.edu/pettyk/FSP/>, Accessed November 20, 2006.
- [2] J.G. Wardrop, Some theoretical aspects of road traffic research, in: Proceedings of the Institution of Civil Engineers, Part II, Vol. I, 1952, pp. 325–362.
- [3] N. Gartner, C.J. Messer, A.K. Rathi, Revised Monograph of Traffic Flow Theory: A State-of-the-Art Report, Transportation Research Board, 2001.
- [4] D. Ni, Determining traffic flow characteristics by definition for application in ITS, IEEE Tran. Intellig. Transport. Syst. 8 (2) (2007) 181–187.
- [5] M. Szeto, D. Gazis, Application of Kalman filtering to the surveillance and control of traffic systems, Transport. Sci. 6 (4) (1972) 419–439.
- [6] L.C. Edie, Discussion on traffic stream measurements and definitions, in: Proceedings of the 2nd Int. Symp. Theory Traffic Flow, Paris, France, 1963, pp. 139–154.
- [7] Y. Makigami, G. Newell, R. Rothery, Three-dimensional representation of traffic flow, Transport. Sci. 5 (3) (1971) 302–313.
- [8] D.R. Drew, Traffic Flow Theory and Control, McGraw-Hill Book Company, 1968, (Chapter 12).
- [9] B. Greenshields, A study of traffic capacity, Proc. Highway Res. Board 14 (1934) 448–477.
- [10] H. Greenberg, An analysis of traffic flow, Operat. Res. 7 (1959) 78–85.
- [11] R. Underwood, Speed, volume and density relationships, quality and theory of traffic flow, Yale University Report, New Haven, CT, 1961.
- [12] J. Drake, J. Schofer, A. May, A statistical analysis of speed density hypotheses, Highway Res. Record 154 (1967) 53–87.
- [13] D. Drew, Deterministic aspects of freeway operations and control, Highway Res. Record 99 (1965) 48–58.
- [14] L. Pipes, Car following models and fundamental diagram of road traffic, Transport. Res. 1 (1967) 21–29.
- [15] P. Munjal, L. Pipes, Propagation of on-ramp density perturbations on uni-directional and two- and three-lane freeways, Transport. Res. B 5 (4) (1971) 241–255.
- [16] L. Edie, Car-following and steady state theory for noncongested traffic, Operat. Res. 9 (1961) 66–76.
- [17] A.D. May, Traffic Flow Fundamentals, Prentice-Hall, Englewood Cliffs, NJ, 1989.
- [18] H. Wang, J. Li, Q.-Y. Chen, D. Ni, Logistic modeling of the equilibrium speed-density relationship, Transport. Res. A 45 (6) (2011) 554–566.
- [19] H. Wang, D. Ni, Q.-Y. Chen, J. Li, Stochastic modeling of equilibrium speed-density relationship, J. Adv. Transport. 47 (1) (2013) 126–150.
- [20] J. Li, H. Wang, Q.-Y. Chen, D. Ni, Analysis of LWR model with fundamental diagram subject to uncertainties, Transportmetrica 8 (6) (2012) 387–405.
- [21] H. Payne, FREFLO: a macroscopic simulation model for freeway traffic, Transport. Res. Record 722 (1979) 68–77.
- [22] R. Illner, C.S. Bohun, S. McCollum, T.V. Roode, Mathematical Modelling: A Case Study Approach, American Mathematical Society, Providence, RI, 2005.
- [23] R. Knobel, An Introduction to the Mathematical Theory of Waves, American Mathematical Society, Providence, RI, 2000.
- [24] M. Lighthill, G. Whitham, On kinematic waves II. A theory of traffic flow on long crowded roads, Proc. Royal Society of London, Part A 229 (1178) (1955) 317–345.

- [25] P. Richards, Shock waves on the highway, *Operat. Res.* 4 (1956) 42–51.
- [26] A. May, FREQ User Manual, Tech. Rep., California Department of Transportation Berkeley, CA, 1998.
- [27] P. Michalopoulos, Dynamic freeway simulation program for personal computers, *Transport. Res. Record* 971 (1984) 68–79.
- [28] C. Daganzo, The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory, *Transport. Res. B* 28 (4) (1994) 269–287.
- [29] C. Daganzo, The cell transmission mode, Part II: Network traffic, *Transport. Res. B* 29 (2) (1995) 79–93.
- [30] G. Newell, A simplified theory on kinematic waves in highway traffic, Part I: General theory, *Transport. Res. B* 27 (4) (1993) 281–287.
- [31] G. Newell, A simplified theory on kinematic waves in highway traffic, Part II: Queueing at freeway bottlenecks, *Transport. Res. B* 27 (4) (1993b) 289–303.
- [32] G. Newell, A simplified theory on kinematic waves in highway traffic, Part III: Multi-destination flows, *Transport. Res. B* 27 (4) (1993c) 305–313.
- [33] B. Son, A study of G.F. Newell's simplified theory of kinematic waves in highway traffic, Ph.D. thesis, Department of Civil Engineering, University of Toronto, Canada, 1996.
- [34] V.F. Hurdle, B. Son, Road test of a freeway model, *Transport. Res. A* 34 (7) (2000) 537–564.
- [35] D. Ni, Extension and generalization of Newell's simplified theory of kinematic waves, Ph.D. thesis, Department of Civil and Environmental Engineering, Georgia Institute of Technology, 2004.
- [36] D. Ni, J.D. Leonard, B.M. Williams, The network kinematic waves model: a simplified approach to network traffic, *J. Intell. Transport. Syst. Technol. Plan. Operat.* 10 (1) (2006) 1–14.
- [37] H. Payne, Models of freeway traffic and control, in: Simulation Council Proceedings, vol. 1, 1971, pp. 51–61.
- [38] G. Whitham, *Linear and Nonlinear Waves*, John Wiley and Sons Inc., New York, 1974.
- [39] W. Phillips, A kinetic model for traffic flow with continuum implications, *Transport. Plan. Technol.* 5 (1979) 131–138.
- [40] R. Kühne, Macroscopic freeway model for dense traffic—stop-start waves and incident detection, in: Ninth International Symposium on Transportation and Traffic Theory, 1984, pp. 20–42.
- [41] R. Kühne, Freeway control and incident detection using a stochastic continuum theory of traffic flow, in: 1st International Conference on Applied Advanced Technology in Transportation Engineering, San Diego, CA, 1989, pp. 287–292.
- [42] B. Kerner, P. Konhäuser, Cluster effect in initially homogeneous traffic flow, *Phys. Rev. E* 48 (4) (1993) 2335–2338.
- [43] P. Michalopoulos, P. Yi, A. Lyrintzis, Continuum modelling of traffic dynamics for congested freeways, *Transport. Res. B* 27 (1993) 315–332.
- [44] H. Zhang, A theory of nonequilibrium traffic flow, *Transport. Res. B* 32 (7) (1998) 485–498.
- [45] M. Treiber, A. Hennecke, D. Helbing, Derivation, properties and simulation of a gas-kinetic-based, nonlocal traffic model, *Phys. Rev. E* 59 (1) (1999) 239–253.
- [46] J. Yi, H. Lin, L. Alvarez, R. Horowitz, Stability of macroscopic traffic flow modeling through wavefront expansion, *Transport. Res. B* 37 (7) (2003) 661–679.
- [47] C. Daganzo, Requiem for second-order fluid approximations of traffic flow, *Transport. Res. B* 29 (4) (1995) 277–286.
- [48] I. Prigogine, F.C. Andrews, A Boltzmann-like approach for traffic flow, *Operat. Res.* 8 (6) (1960) 789–797.

- [49] J.M. del Castillo, P. Pintado, F.G. Benitez, A formulation for the reaction time of traffic flow models, in: Proceedings of the 12th Int. Symp. Theory Traffic Flow, Paris, France, 1993, pp. 387–405.
- [50] P.G. Michalopoulos, J.K. Lin, D.E. Beskos, Integrated modelling and numerical treatment of freeway flow, *Appl. Math. Model.* 11 (401) (1987) 447–458.
- [51] C.J. Leo, R.L. Pretty, Numerical simulation of macroscopic continuum traffic models, *Transport. Res. B* 26 (3) (1992) 207–220.
- [52] L.A. Pipes, An operational analysis of traffic dynamics, *J. Appl. Phys.* 24 (1953) 271–281.
- [53] T. Forbes, Human factor considerations in traffic flow theory, *Highway Res. Record* 15 (1963) 60–66.
- [54] T. Forbes, M. Simpson, Driver and vehicle response in freeway deceleration waves, *Transport. Sci.* 2 (1) (1968) 77–104.
- [55] R. Chandler, R. Herman, E. Montroll, Traffic dynamics: studies in car following, *Operat. Res.* 6 (1958) 165–184.
- [56] D.C. Gazis, R. Herman, R. Rothery, Non-linear follow the leader models of traffic flow, *Operat. Res.* 9 (1961) 545–567.
- [57] P. Gipps, A behavioral car following model for computer simulation, *Transport. Res. B* 15 (1981) 105–111.
- [58] G.F. Newell, Nonlinear effects in the dynamics of car following, *Operat. Res.* 9 (2) (1961) 209–229.
- [59] G. Newell, A simple car-following theory: a lower order model, *Transport. Res. B* 36 (3) (2002) 195–205.
- [60] M. Treiber, A. Hennecke, D. Helbing, Congested traffic states in empirical observations and microscopic simulations, *Phys. Rev. E* 62 (2000) 1805–1824.
- [61] D. Helbing, A. Hennecke, V. Shvetsov, M. Treiber, Micro- and macro-simulation of freeway traffic, *Math. Comput. Model.* 35 (5) (2002) 517–547.
- [62] M. Van Aerde, Single regime speed-flow-density relationship for congested and uncongested highways, in: Presented at the 74th Transportation Research Board (TRB) Annual Meeting, Paper number 950802, Washington, DC, 1995.
- [63] M. Van Aerde, H. Rakha, Multivariate calibration of single regime speed-flow-density relationships, in: Proceedings of the 6th 1995 Vehicle Navigation and Information Systems Conference, Seattle, WA, USA, 1995, pp. 334–341.
- [64] R. Wiedemann, Simulation des Straenverkehrsflusses, Ph.D. thesis, Schriftenreihe des Instituts für Verkehrswesen der Universität Karlsruhe, Germany, 1974.
- [65] R. Benekohal, J. Treiterer, Carsim: car following model for simulation of traffic in normal and stop-and-go conditions, *Transport. Res. Record* 1194 (1988) 99–111.
- [66] T. Forbes, H. Zagorski, E. Holshouser, W.A. Deterline, Measurement of driver reaction to tunnel conditions, *Proc. Highway Res. Board* 37 (1958) 345–357.
- [67] I. Kosonen, HUTSIM—urban traffic simulation and control model: principles and applications, Ph.D. thesis, Helsinki University of Technology, 1999.
- [68] H. Jia, Z. Juan, A. Ni, Develop a car-following model using data collected by “five-wheel system”, in: The Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems, vol. 1, 2003, pp. 346–351.
- [69] S. Panwai, H. Dia, Neural agent car-following models, *IEEE Trans. Intell. Transport. Syst.* 8 (1) (2007) 60–70.
- [70] D. Ni, Challenges and strategies of transportation modeling and simulation under extreme conditions, *Int. J. Emergen. Manage.* 3 (4) (2006) 298–312.
- [71] D. Ni, A framework for new generation transportation simulation, in: Proceedings of Winter Simulation Conference '06, Portola Plaza Hotel, Monterey, CA, December 3–6, 2006.

- [72] D. Ni, D. Henclewood, Simple engine models for VII-enabled in-vehicle applications, *IEEE Trans. Vehicular Technol.* 57 (5) (2008) 2695–2702.
- [73] B.K. Powell, A dynamic model for automotive engine control analysis, in: 18th IEEE Decision and Control Conference, 1979.
- [74] D.J. Dobner, A mathematical engine model for development of dynamic engine control, Tech. Rep., Society of Automotive Engineers (SAE) Report No. 800054, 1980.
- [75] D.J. Dobner, Dynamic engine models for control development. Part I: Nonlinear and linear model formulation, *Int. J. Vehicle Design* SP4 (1983) 54–74.
- [76] F.E. Coates, R.D. Fruechte, Dynamic engine models for control development. Part II: Application to idle speed control, *Int. J. Vehicle Design* SP4 (1983) 75.
- [77] J.D. Powell, A review of IC engine models for control system design, in: Proceedings of 10th World Congress on Automatic Control, Munich, Germany, 1987.
- [78] J. Moskwa, J. Hedrick, Automotive engine modeling for real time control application, in: Proceedings of the 1987 American Control Conference, Minneapolis, MN, 1987, pp. 341–346.
- [79] P. Yoon, M. Sunwoo, A nonlinear dynamic modelling of SI engines for controller design, *Int. J. Vehicle Design* 26 (2001) 277–297.
- [80] D. Cho, J.K. Hedrick, Automotive powertrain modeling for control, *Trans. ASME J. Dyn. Syst. Measur. Control* 111 (1989) 568–576.
- [81] J.J. Moskwa, Automotive engine modeling for real time control, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1988.
- [82] B. Akinci, C. Hendrickson, I. Karaesmen, Exploiting motor vehicle information and communications technology for transportation engineering, *J. Transport. Eng.* 129 (2003) 469–474.
- [83] P.F. Puleston, G. Monsees, S.K. Spurgeon, Air-fuel ratio and speed control for low emission vehicles based on sliding mode techniques, *Proceedings of the Institution of Mechanical Engineers. Part I: Journal of Systems and Control Engineering* 216 (2002) 117–124.
- [84] G. Rizzoni, Estimate of indicated torque from crankshaft speed fluctuations: a model for the dynamics of the IC engine, *IEEE Trans. Vehicular Technol.* 38 (1989) 168–179.
- [85] G. Rizzoni, Stochastic model for the indicated pressure process and the dynamics of the internal combustion engine, *IEEE Trans. Vehicular Technol.* 38 (1989) 180–192.
- [86] P.R. Crossley, J.A. Cook, A nonlinear engine model for drivetrain system development, in: International Conference on Control, Vol. 2, Edinburgh, UK, 1991, pp. 921–925.
- [87] C. Hong, Automotive dynamic performance simulator for vehicular powertrain system design, *Int. J. Vehicle Design* 16 (1995) 264–281.
- [88] J.B. Heywood, Internal combustion engine fundamentals, McGraw-Hill, New York, 1989.
- [89] C.S. Daw, M.B. Kennel, C.E.A. Finney, F.T. Connolly, Observing and modeling nonlinear dynamics in an internal combustion engine, *J. Amer. Phys. Soc. Rev. E* 57 (1998) 2811–2819.
- [90] Y. Shiao, C.H. Pan, J.J. Moskwa, Advanced dynamic spark ignition engine modelling for diagnostics and control, *Int. J. Vehicle Design* 15 (1994) 578–596.
- [91] O. Chiavola, Integrated modelling of internal combustion engine intake and exhaust systems, *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy* 215 (2001) 495–506.
- [92] Z.D. Ma, N.C. Perkins, An efficient multibody dynamics model for internal combustion engine systems, *Multibody Syst. Dyn.* 10 (2003) 363–391.
- [93] M. Kabganiyan, R. Kazemi, A new strategy for traction control in turning via engine modeling, *IEEE Trans. Vehicular Technol.* 50 (2001) 1540–1548.

- [94] J.R. Wagner, D.M. Dawson, L. Zeyu, Nonlinear air-to-fuel ratio and engine speed control for hybrid vehicles, *IEEE Trans. Vehicular Technol.* 52 (2003) 184–195.
- [95] S. Delprat, J. Lauber, T.M. Guerra, J. Rimaux, Control of a parallel hybrid powertrain: optimal control, *IEEE Trans. Vehicular Technol.* 53 (2004) 872–881.
- [96] J.J. Scillieri, J.H. Buckland, J.S. Freudenberg, Reference feedforward in the idle speed control of a direct-injection spark-ignition engine, *IEEE Trans. Vehicular Technol.* 54 (2005) 51–61.
- [97] K.L. Butler, M. Ehsani, P. Kamath, Matlab-based modeling and simulation package for electric and hybrid electric vehicle design, *IEEE Trans. Vehicular Technol.* 48 (1999) 1770–1778.
- [98] D.W. Gao, C. Mi, A. Emadi, Modeling and simulation of electric and hybrid vehicles, in: *Proc. IEEE*, 95 (2007) 729–745.
- [99] J.A. Cook, B.K. Powell, Modeling of an internal combustion engine for control analysis, *IEEE Control Syst. Mag.* 8 (4) (1988) 20–26.
- [100] D. Swaroop, J.K. Hedrick, C.C. Chien, P. Ioannou, Comparison of spacing and headway control laws for automatically controlled vehicles, *Vehicle Syst. Dyn.* 23 (1994) 597–625.
- [101] D.J. Grieve, Simulation of dynamic systems—MECH 337, <http://www.tech.plym.ac.uk/sme/mech331/partb1.htm>. Accessed November 20, 2006.
- [102] G. Genta, *Motor Vehicle Dynamics: Modeling and Simulation*, World Scientific, Singapore, 2003.
- [103] M.D. Artamonov, V.A. Ilarionov, M.M. Morin, *Motor Vehicles: Fundamentals and Design*, Mir Publishers, Moscow, 1976.
- [104] G. Bar-Meir, *Fundamentals of compressible fluid mechanics*, Potto Project, 2007.
- [105] M. Ehsani, Y.M. Gao, S.E. Gay, A. Emadi, *Modern Electric, Hybrid Electric, and Fuel Cell Vehicles*, CRC Press, Boca Raton, FL, 2005.
- [106] D. Henclewood, D. Ni, A dynamic-interactive-vehicle model for modeling traffic beyond the microscopic level, *Int. J. Vehicle Informat. Commun. Syst.* 2 (1–2) (2009).
- [107] R.G. Mortimer, L. Segel, H. Dugoff, J.D. Campbell, C.M. Jorgeson, R.W. Murphy, Brake force requirement study: driver-vehicle braking performance as a function of brake system design variables, Report Number: DOT/HS 800 253, Tech. Rep., Highway Safety Research Institute Ann Arbor, MI, 1970.
- [108] T.D. Gillespie, *Fundamentals of Vehicle Dynamics*, Society of Automotive Engineers, Inc., Troy, MI, 1992.
- [109] E.N. Moret, Dynamic modeling and control of a car-like robot, Master's thesis, Department of Electrical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, 2003.
- [110] D. Ni, A unified perspective on traffic flow theory, Part I: The field theory, *Appl. Math. Sci.* 7 (39) (2013) 1929–1946.
- [111] R.M. Michaels, Perceptual factors in car following, in: *Proceedings of the 2nd International Symposium on the Theory of Road Traffic Flow* (London, England), OECD, 1963.
- [112] S. Kikuchi, P. Chakroborty, Car-following model based on a fuzzy inference system, *Transport. Res. Record* 1365 (1992) 82–91.
- [113] D. Ni, J.D. Leonard, C. Jia, J. Wang, Vehicle longitudinal control and traffic stream modeling, *Transport. Sci.*, 2015, Published online July 10, 2015. Permalink: <http://dx.doi.org/10.1287/trsc.2015.0614>.
- [114] D. Ni, Multiscale modeling of traffic flow, *Math. Aeterna* 1 (01) (2011) 27–54.
- [115] D. Ni, A unified perspective on traffic flow theory, Part I: The field theory, *Appl. Math. Sci.* 7 (39) (2013) 1929–1946.
- [116] M. Treiber, A. Kesting, *Traffic Flow Dynamics*, Springer, New York, 2013.

- [117] D. Ni, H. Wang, A unified perspective on traffic flow theory, Part III: Validation and benchmarking, *Appl. Math. Sci.* 7 (39) (2013) 1929–1946.
- [118] J.M. del Castillo, F.G. Benítez, On the functional form of the speed-density relationship—I: General theory, *Transport. Res. B Meth.* 29 (5) (1995) 373–389.
- [119] H. Rakha, Validation of Van Aerde's simplified steady-state car-following and traffic stream model, *Transport. Lett. Int. J. Transport. Res.* 1 (3) (2009) 227–244.
- [120] J.M. Del Castillo, Three new models for the flow-density relationship: derivation and testing for freeway and urban data, *Transportmetrica* 8 (6) (2012) 443–465.
- [121] M. Koshi, M. Iwasaki, I. Ohkura, Some findings and an overview on vehicular flow characteristics, in: 8th International Symposium on Transportation and Traffic Flow Theory, Toronto, 1983, pp. 403–426.
- [122] J.H. Banks, Freeway speed-flow-concentration relationships: more evidence and interpretations, *Transport. Res. Record* 1225 (1989) 53–60.
- [123] H. Rakha, M. Arafah, Calibrating steady-state traffic stream and car-following models using loop detector data, *Transport. Sci.* 44 (2) (2010) 151–168.
- [124] Highway Capacity Manual, Transportation Research Board, 2010.
- [125] W.J.M. Rankine, On the thermodynamic theory of waves of finite longitudinal disturbances, *Philosophical Transactions of the Royal Society* 160 (1870) 277–288.
- [126] H. Hugoniot, Propagation des mouvements dans les corps et spécialement dans les gaz parfaits" (in French), *Journal de l'Ecole Polytechnique* 57 (1887) 3.
- [127] D. Ni, A unified perspective on traffic flow theory, Part II: The unified diagram, *Appl. Math. Sci.* 7 (39) (2013) 1929–1946.
- [128] Highway Capacity Manual, Bureau of Public Roads, U.S. Dept. of Commerce, 1950.
- [129] D. Ni, A unified perspective on traffic flow theory, Part II: The unified diagram, *Appl. Math. Sci.* 7 (40) (2013) 1947–1963.
- [130] H. Wang, J. Li, Q.-Y. Chen, D. Ni, Representing the fundamental diagram: the pursuit of mathematical elegance and empirical accuracy, in: Pre-print CD-ROM, the 89th Transportation Research Board (TRB) Annual Meeting, Paper number 10-1354, Washington, D.C., 2010.
- [131] J.M. del Castillo, F.G. Benítez, On the functional form of the speed-density relationship—II: Empirical investigation, *Transport. Res. B Meth.* 29 (5) (1995) 391–406.
- [132] M. Abe, Theoretical analysis on vehicle cornering behaviours in braking and in acceleration, *Vehicle Syst. Dyn.* 14 (1-3) (1985) 140–143.
- [133] E.M. Lowndes, J. David, Development of an intermediate degree of freedom vehicle dynamics model for optimal design studies, *Amer. Soc. Mech. Eng. DE* 106 (2000) 19–24.
- [134] W. Wierwille, G. Gagne, J. Knight, An experimental study of human operator models and closed-loop analysis methods for high-speed automobile driving, *IEEE Trans. Human Factors Electron. HFE-8* (3) (1967) 187–201.
- [135] T. Gordon, M. Best, On the synthesis of driver inputs for the simulation of closed-loop handling manoeuvres, *Int. J. Vehicle Design* 40 (1-3) (2006) 52–76.
- [136] U. Kramer, G. Rohr, A model of driver behaviour, *Ergonomics* 25 (10) (1982) 891–907.
- [137] Z. Gao, N. Zheng, H. Guan, K. Guo, Application of driver direction control model in intelligent vehicle's decision and control algorithm, in: *IEEE Trans. Intelligent Vehicle Symposium*, 2002, vol. 2, 2002, pp. 413–418.
- [138] C.C. Macadam, G.E. Johnson, Application of elementary neural networks and preview sensors for representing driver steering control behaviour, *Vehicle Syst. Dyn.* 25 (1) (1996) 3–30.

- [139] Y. Lin, P. Tang, W. Zhang, Q. Yu, Artificial neural network modelling of driver handling behaviour in a driver-vehicle-environment system, *Int. J. Vehicle Design* 37 (1) (2005) 24–45.
- [140] A. Zadeh, A. Fahim, M. El-Gindy, Neural network and fuzzy logic applications to vehicle systems: literature survey, *Int. J. Vehicle Design* 18 (2) (1997) 132–193.
- [141] M. Irmscher, T. Jurgensohn, H.-P. Willumeit, Driver models in vehicle development, *Vehicle Syst. Dyn.* 33 (Suppl) (1999) 83–93.
- [142] C.C. Macadam, Understanding and modeling the human driver, *Vehicle Syst. Dyn.* 40 (1–3) (2003) 101–134.
- [143] P. Gipps, A model for the structure of lane-changing decisions, *Transport. Res. B* 20 (1986) 403–414.
- [144] K. Ahmed, M. Ben-Akiva, H. Koutsopoulos, R. Mishalani, Models of freeway lane-changing and gap acceptance behavior, in: Proceedings of the 13th International Symposium on the Theory of Traffic Flow and Transportation, 1996, pp. 501–515.
- [145] P. Hidas, K. Behbahanizadeh, Microscopic simulation of lane changing under incident conditions, in: Proceedings of the 14th International Symposium on the Theory of Traffic Flow and Transportation, 1999, pp. 53–69.
- [146] Y. Zhang, L. Owen, J. Clark, A multi-regime approach for microscopic traffic simulation, in: 77th Transportation Research Board Annual Meeting, 1998.
- [147] R. Sukthankar, Situation awareness for tactical driving, Ph.D. thesis, Carnegie Mellon University, 1997.
- [148] M. Raff, J. Hart, A volume warrant for urban stop signs, tech. rep., Eno Foundation for Highway Traffic Control, Saugatuck, Connecticut, 1950.
- [149] S.M. Velan, M. Van Aerde, Gap acceptance and approach capacity at unsignalized intersections, *ITE J.* 66 (3) (1996) 40–45.
- [150] M. Hamed, S. Sama, R. Batayneh, Disaggregate gap-acceptance model for unsignalised t-intersections, *J. Transport. Eng. ASCE* 123 (1) (1997) 36–42.
- [151] R. Herman, G. Weiss, Comments on the highway crossing problem, *Operat. Res.* 9 (1961) 838–840.
- [152] D. Drew, L. LaMotte, J. Buhr, J. Wattleworth, Gap acceptance in the freeway merging process, Tech. Rep., Texas Transportation Institute, 1967, pp. 430–432.
- [153] R. Hewitt, Using probit analysis with gap acceptance data, Tech. Rep., Department of Civil Engineering, University of Glasgow 1992.
- [154] R. Rossi, C. Meneguzzi, The effect of crisp variables on fuzzy models of gap-acceptance behaviour, in: Proceedings of the 13th Mini-EURO Conference: Handling Uncertainty in the Analysis of Traffic and Transportation Systems, 2002.
- [155] S. Hoogendoorn, P. Bovy, State-of-the-art of vehicular traffic flow modelling, *Proceedings of the IMechE Part I, J. Syst. Control Eng.* 215 (4) (2001) 283–303.
- [156] L. Smith, R. Beckman, D. Anson, K. Nagel, M. Williams, TRANSIMS: transportation analysis and simulation system, in: Fifth National Conference on Transportation Planning Methods Applications, vol. II, 1995.
- [157] M.E. Ben-Akiva, M. Bierlaire, H. Koutsopoulos, R. Mishalani, DynaMIT: a simulation-based system for traffic prediction, in: DACCORS Short Term Forecasting Workshop, 1998.
- [158] G.-L. Chang, T. Junchaya, A.J. Santiago, A real-time network traffic simulation model for atm applications: part i—simulation methodologies, *J. Intell. Transport. Syst.* 1 (3) (1994) 227–241.
- [159] I. Prigogine, *A Boltzmann-like Approach to the Statistical Theory of Traffic Flow, Theory of traffic Flow*, Elsevier, Amsterdam, 1961.
- [160] F. Haight, Vehicles as particles. (Book Reviews: Kinetic Theory of Vehicular Traffic by Prigogine and Herman), *Science* 173 (3996) (1971) 513.

- [161] S. Paveri-Fontana, On Boltzmann-like treatments for traffic flow: a critical review of the basic model and an alternative proposal for dilute traffic analysis, *Transport. Res.* 9 (1975) 225–235.
- [162] D. Helbing, Theoretical foundation of macroscopic traffic models, *Phys. A Stat. Theoret. Phys.* 219 (3-4) (1995) 375–390.
- [163] W. Phillips, Kinetic model for traffic flow, *Tech. Rep., Report DOT/RSPD/DPB/50-77/17*. U. S. Department of Transportation, 1977.
- [164] P. Nelson, A kinetic model of vehicular traffic and its associated bimodal equilibrium solutions, *Transport Theory Stat. Phys.* 24 (1995) 383–409.
- [165] R. Wegener, A. Klar, A kinetic model for vehicular traffic derived from a stochastic microscopic model, *Transport Theory Stat. Phys.* 25 (1996) 785–798.
- [166] A. Klar, R. Wegener, A hierarchy of models for multilane vehicular traffic (Part I: Modeling and Part II: Numerical and stochastic investigations), *SIAM J. Appl. Math.* 59 (1999) 983–1011.
- [167] D. Helbing, Traffic and related self-driven many-particle systems, *Rev. Mod. Phys.* 73 (2001) 1067–1141.
- [168] S. Yager, CORQ – a model for predicting flows and queues in a road corridor, *Transport. Res. Record* 533 (1975) 77–87.
- [169] R.C. Tolman, *The Principles of Statistical Mechanics*, Dover Publications, New York, 1980.
- [170] S. Harris, *An Introduction to the Theory of the Boltzmann Equation*, Dover Books on Physics, Dover Publications, New York, 2004.
- [171] C. Daganzo, A finite difference approximation of the kinematic wave model of traffic flow, *Transport. Res. B* 29 (4) (1995) 261–276.

INDEX

Note: Page numbers followed by *f* indicate figures and *t* indicate tables.

A

- Ackerman angle, 279
- Acoustic/ultrasonic sensors
 - advantages, 10
 - data collection, 10
 - disadvantages, 10
 - parking sensor, 9-10
 - pedestrian/vehicle and sensor distance, 9-10
- Aerial/satellite imaging
 - advantages, 11
 - data collection, 11
 - disadvantages, 11
 - manned/unmanned helicopters, 10-11, 11*f*
 - vehicle counts, vehicle speeds and traffic density, 10-11
- AIMSUN, 245-247

B

- Bernoulli model
 - compressible fluid mechanics, 267
 - empirical comparison, 268
 - fluid velocity, 265-266
 - ideal gas law, 266
 - principle, 262, 265-266
- Boltzmann equation
 - conservation law, 370
 - initial and boundary conditions, 370-371
 - longitudinal control model, 370

C

- California Vehicle Code, 183, 185-186
- Car-following models
 - CARSIM model, 239-240
 - Forbes model, 187-188
 - macroscopic benchmarking, 191-193
 - microscopic benchmarking, 188-191
 - microscopic models, taxonomy of, 246*f*

neural network model

- artificial neural networks, 241-242

back-propagation neural network, 242

data collection, 243

driver's operational control, 242

first input-output pattern, 243

neurons, 241-242

training process, 243

Pipes model, 183-187

psychophysical model (*see* VISSIM)

driver's operational control, 237

driving regimes, 239

illustration of, 238*f*

operating point, 239

psychological activities and physical behavior, 239

vehicles collision, 237

rule-based model, 240-241

Cell transmission model (CTM)

divergence scenario, 146-147

mainline scenario, 140-142

merger scenario

demand exceeds supply, both upstream links congestion, 145

demand exceeds supply, one upstream link congestion, 144

freeway merge, 142*f*

queuing model, 142

supply exceeds demand, 143

minimum principle, 139-140

taxonomy of, 170*f*

1964 Chevrolet Corvair, 268-269

Connected vehicle technology (CVT), 261-262

Conservation law

continuity equation

finite difference, derivation I, 72-74

finite difference, derivation II, 74-75

fluid dynamics, derivation III, 75-77

roadway traffic, 72

- Conservation law (*Continued*)

scalar conservation law, derivation IV, 77–78

three-dimensional representation of traffic flow, derivation V, 78

first-order dynamic model, 79–80

flow-speed-density relationship/identity, 71

mass conservation, 72

one-dimensional compressible fluid, 72

pairwise relationships/equilibrium models, 71

PDE, 92

traffic flow theory, 71–72

traffic states, 72
- Continuum flow models

Kerner and Konhäuser's model, 166

LWR model, 166

Michalopoulos et al. model, 166

Phillips's model, 166

PW model, 166

relative merits of, 167–168

Treiber et al. model, 167

Zhang's model, 166
- Control theory, 360, 363

CORQ, 362

CORSIM, 245–247

Coulomb's law, 207, 287–288, 341–342

CTM. *See* Cell transmission model (CTM)

CVT. *See* Connected vehicle technology (CVT)
- D**

d'Alembert solution, 88–89

Del Castillo model, 332

Desired measure model, 245–247

Deterministic models, 360–361

Discrete choice model

lane-changing model, 175

lateral direction, 178

route-choice model, 173–175

Drake et al. model, 203, 205, 332

Drew model, 204, 205

Dynamic interactive vehicle (DIV) model

automotive engineering, 275

brake pedal position, 275–276

calibration and validation, 279–280
- high computational efficiency and modeling fidelity, 275

modeling lateral movement, 278–279

modeling longitudinal movement

aerodynamic drag, 277

braking performance, 277

grade resistance, 278

modeling acceleration performance, 276

Newton's second law of motion, 276

steering angle, 275–276

throttle position, 275–276

DynaMIT, 361–362

DYNASMART, 361–362
- E**

Edie model, 59–62

Energy conservation, 164, 168, 371–372

Engine modeling

accessibility, 262

accuracy, 262

calibration, 262

computational efficiency, 262

cross-comparison of, 273

CVT, 261–262

existing engine models

direct-injection spark-ignition engine model, 263

filling and emptying method, 263

hybrid vehicle modeling, 263

IC engines, 262–263

linear engine model, 263–264

spark-ignition engine, 263

teaching purpose, 263–264

formulation, 262

internal combustion engines, 261–262

mathematical engine models

Bernoulli model, model III, 265–268

parabolic model, model II, 265

polynomial model, model I, 264

parameter values, 273

validation and comparison of

automotive engines, 268–269

1964 Chevrolet Corvair engine, 268–269, 271*f*

empirical engine power and torque curves, 268–269

- 2006 Honda Civic engine, 268–269,
270f
 MAPE, 271–272
 2008 Mercedes CLS engine, 268–269,
270f
 model I, 269
 model II, 269
 model III, 269
 2006 Pagani Zonda engine, 268–269,
270f
 technical specifications, 269t
- Equilibrium traffic flow models**
 multiregime models, 59–62, 62*t*, 64*f*
 single-regime models
 comparison of, 61*f*
 equilibrium/steady-state relationship, 51–52
 flow, speed and density, 52–55
 free-flow speed, 58–59
 Georgia NaviGAtor, 51
 Greenshields model, 55–58
 one-equation models, 59
 point sensor, 51, 53*f*
 traffic speed, 51–52
 video images, 51
 speed-density relationship
 observed mean and standard deviation, 62–63, 65*f*
 scattering effect, 62–63
 three-dimensional representation, 65*f*
- F**
- Federal Motor Carrier Safety Administration, 280
- Field theory
 congestion, 301–302
 driver's interaction, 294–295
 driver's responsiveness, 292–293
 driver's strategy, 295–296
 existing knowledge base
 engineering disciplines, 302–303
 physical science, 304
 traffic flow theories, 302
 mobility, 301
 motivation, 283–284
 physical field, 291–292
 simplification of
- decouples equations, 297
 directional response, 297–300
 forces, 297
 generic theory, 296
 lateral direction, 299–300
 longitudinal direction, 297–299
 Occam's razor, 296
 traffic flow, physical basis of
 electromagnetic phenomena, 287–288
 mechanics phenomena, 285–287
 physical laws and social rules, 284
 statistical mechanics phenomena, 289–291
 three systems, 285*f*
 transportation system, 284
 wave phenomena, 288–289
- Forbes model
 applications (*see* Pipes model)
 macroscopic benchmarking, 191–193
 perception-reaction time, 187–188
 properties (*see* Pipes model)
 safety rule, 187
 steady-state model, 331–332
- FREFOLO
 numerical solutions
 conservation law, 135
 dynamic speed-density relationship, 136
 equilibrium speed-density relationship, 137
 macroscopic traffic simulation model, 135
 PW model, 163–164
- FREQ, 137, 362
- G**
- Gap-acceptance models, 245, 360–361
- General Motors (GM) models
 development of
 driver's control maneuver, 195
 GM1, 196
 GM2, 196–197
 GM3, 197
 GM4, 197–198
 GM5, 198
 stimuli and sensitivity, 195–196

General Motors (GM) models (*Continued*)
 limitations of
 intimate pair, 208–209
 mechanism of motion, attraction, 207
 slow start, 208, 208*f*
 universal car following, 207
 macroscopic benchmarking, 205
 microscopic benchmarking, 198–199
 microscopic-macroscopic bridge
 Drake et al. model (Northwestern Model), 203
 Drew model, 204
 GM5, 201
 Greenberg model, 201–202
 Greenshields model, 203
 macroscopic equilibrium models, 204
 microscopic car-following models, 204
 Pipes and Forbes models, 204
 Pipes-Munjal model, 203
 single-regime models, 201*t*
 Underwood model, 203
 traffic flow theory, 195
 Wilson, Charles, 195
 Georgia State Route 400 (GA400) data, 13–15
 Gipps model
 fundamental diagram, 219*f*
 macroscopic benchmarking, 218
 microscopic benchmarking, 216–218, 217*f*
 model formulation
 car-following scenario, 212*f*
 Forbes rule, 211
 free-flow regime, 214
 perception-reaction process, 212
 Pipes rule, 211
 quadratic inequality, 213
 safe driving rule, 211
 speed choice model, 214
 properties of, 215–216
 steady-state model, 331–332
 Global positioning system (GPS)
 advantages, 9
 automotive navigation and traffic time studies, 8
 cell phones, 8
 data collection, 9

disadvantages, 9
 satellite-based navigation system, 8–9
 spatial-temporal points, 9
 triangulation, 8–9
 vehicle's trajectory, 19–21, 21*f*
 Gradient catastrophe, 96
 Greenberg model, 201–202, 205, 332
 Greenshields model
 dynamic traffic flow modeling, 58
 flow-density relationship, 56*f*
 flow-density-speed relationship, 57*f*
 fundamental diagram, 57–58, 60*f*
 highway capacity and LOS, 58
 LWR model, 115–116
 microscopic-macroscopic bridge, 203
 speed-density relationship, 55, 55*f*
 speed-flow relationship, 56*f*, 57
 traffic flow theory, 58
 vehicular traffic flow, 57–58

H

High-order models
 continuum flow models, 166–167
 first-order partial differential equation, 163
 Kerner and Konhäuser's model (1993), 164–165
 Kühne's model (1984), 164
 macroscopic models, taxonomy of, 168–169, 170*f*
 mass/vehicle conservation, 163
 Michalopoulos et al. (1993) model, 165
 Phillips's model (1979), 164
 PW model (1971), 163–164
 Treiber et al. (1999) model, 165–166
 Zhang's model (1998), 165
 Highway Capacity Manual (HCM), 37
 2006 Honda Civic, 268–269
 HUTSIM, 245–247

I

IDM. *See* Intelligent driver model (IDM)
 IF-THEN rules, 244, 245, 360–361
 Inductive-loop detectors
 actuated signal controller, 4–5
 advantages, 5
 data collection, 5

- disadvantages, 6
 intersections, 3
 red-light-running camera, 4-5
 resonant frequency, 4
I
 Integration, 245-247
 Intelligent driver model (IDM)
 desired spacing, 227
 fundamental diagram, 232*f*
 macroscopic benchmarking, 230, 230*t*
 microscopic benchmarking, 228, 229*f*,
 230*t*
 properties of, 227-228
 Intelligent transportation systems (ITS), 3
 Internal combustion (IC) engines, 261-262
- K**
 Kalman filter technique, 33
 Kinematic waves model, 362
 Kosonen model, 245, 347*f*
 KRONOS
 first-order models, 362
 numerical solutions
 computerized macroscopic traffic simulation model, 137
 discretization, 138*f*
 Greenshields model, 139
 ramp flows and lane changes, 137
- L**
 Lane-changing models, 245, 360-361
 LCM. *See* Longitudinal control model (LCM)
 Lennard-Jones potential, 302-303, 303*f*
 Level of service (LOS), 58
 Lighthill, Whitham and Richards (LWR) model
 conservation law, 114-115
 dynamic traffic flow model, 113
 flow-density relationship
 free-flow, 122
 kinematic wave, 122
 moving observer, 122-123, 124*f*
 shock wave, 122
 traffic, 122
 general $q-k$ relationship, 118-120
 Greenshields model, 115-116
 RIEMANN problem, 117-118
- shock path and queue tail, 120-121
 shock wave solution
 characteristics method, 117
 fast platoon, 117
 kinematic wave, 116
 shock path, 116-117
 slowing down effect, 116-117
 slow platoon, 117
 traffic flow problems
 bottleneck, varying traffic demand,
 123-126, 124*t*, 125*f*
 moving bottleneck, 126, 126*t*, 127*f*
- Lincoln tunnel, 80
 Longitudinal control model (LCM)
 aggressiveness, 332
 applications
 commercial simulation packages, 325
 freeway, 326-327
 macroscopic approach, graphical solution, 327-329
 microscopic approach, deterministic simulation, 329-330
 microscopic approach, random simulation, 330-331
 service analysis procedure, 325
 traffic phenomena, 325
 transportation planners, 325
- desired spacing, 332
 driving force, 308-309
 empirical results
 Amsterdam data, 321*f*
 autobahn data, 316-324, 318*f*
 cloud density, 317
 congested regime, 316-320
 facility types, 322*t*
 free-flow speed and capacity, 316-320
 GA400 data, 316*f*
 Highway 401 data, 316-325, 320*f*
 Interstate 4 data, 316-325, 317*f*
 Newell model, 316-320
 optimized model, 319
 PeMS data, 316-324, 319*f*
 reverse-lambda shape, 316-320
 traffic flow data, 316
 traffic stream models, 316
 traffic stream models, facility types,
 323*t*

- Longitudinal control model (LCM)
(Continued)
- Underwood model, 316–320
 - Forbes model, 331–332
 - General motors models, 331–332
 - Gipps car-following model, 331–332
 - intelligent driver model, 331–332
 - macroscopic model, 310–311
 - microscopic model, 309–310
 - model properties
 - boundary conditions, 313–314
 - consistent modeling approach, 312–313
 - field theory and Newton's second law of motion, 311
 - model flexibility, 314–316
 - one-equation model, 311
 - Newell nonlinear car-following model, 331–332
 - Newton's second law of motion, 308–309
 - Pipes model, 331–332
 - single-regime category, 332
 - steady-state model, 331–332
 - Van Aerde car-following model, 331–332
 - vehicle operational control, 308–309
- M**
- MAPE. *See* Mean absolute percentage error (MAPE)
- Mass conservation, 164
- Mathematical engine models
- Bernoulli model, 265–268
 - parabolic model, 265
 - polynomial model, 264
- MATLAB, 329
- Mean absolute percentage error (MAPE), 271–272
- 2008 Mercedes CLS, 268–269
- Microscopic modeling
- macroscopic benchmarking, 180–182
 - microscopic benchmarking
 - car-following models, 178–179
 - hypothetical driving regimes, 178
 - single driving process, 179–180
 - trial and error, 180
 - modeling scope and time frame
 - car-following decision, 175, 177f
 - categories, 175
 - driver's options, 173
 - driving decisions, 173
 - gap-acceptance decision, 175
 - lane-changing decision, 175
 - operational decision, 175, 177f
 - route-choice decision, 173–175
 - strategic level decision, 174f
 - tactical decision, 175, 176f
 - notation, 175–177
- Modified Greenberg model, 59–62
- Momentum conservation, 164
- Multiscale traffic flow modeling
- consistent multiscale models, 358–359
 - microscopic modeling
 - lateral control model, 368–369
 - longitudinal control, 366–368
 - macroscopic modeling, 371–372
 - mesoscopic modeling, 369–371
 - modeling scales, spectrum of
 - issues of, 362–363
 - macroscopic scale, 362
 - mesoscopic scale, 361–362
 - microscopic scale, 360–361
 - picoscopic scale, 360
 - picoscopic modeling
 - component fields, 364
 - driver-vehicle-environment
 - closed-loop control system, 363
 - driving strategies, 364
 - perceived field, illustration of, 365f
 - traffic simulation, 357–358
 - zoomable simulation, 357–358
- N**
- Navier-Stokes equation of motion, 163–164
- NaviGAtor system, 13–14
- Neural network model
- artificial neural networks, 241–242
 - back-propagation neural network, 242
 - data collection, 243
 - driver's operational control, 242
 - first input–output pattern, 243
 - neurons, 241–242
 - training process, 243
- Neuro-fuzzy hybrid models, 360–361
- Newell model, 316–320, 332

- Newell nonlinear car-following model, 331–332
- Newell nonlinear model
- fundamental diagram, 225 f
 - macroscopic benchmarking, 224, 226 t
 - microscopic benchmarking, 222–224, 223 f , 224 t
- Newell nonlinear car-following model, 221
- properties of, 221–222
- Newell simplified model, 226–227
- Newton's law of universal gravitation, 207
- Next Generation Simulation (NGSIM), 15–17, 245
- Northwestern model, 203, 205
- Numerical solutions
- CTM
 - divergence scenario, 146–147
 - mainline scenario, 140–142
 - merger scenario, 142–146
 - minimum principle, 139–140
 - discretization scheme
 - cell storage, 135
 - free-flow speed, 134
 - time and space, 133, 134 f
- FREFO
- conservation law, 135
 - dynamic speed-density relationship, 136
 - equilibrium speed-density relationship, 137
 - macroscopic traffic simulation model, 135
- FREQ, 137
- KRONOS
- computerized macroscopic traffic simulation model, 137
 - discretization, 138 f
 - Greenshields model, 139
 - ramp flows and lane changes, 137
- P**
- 2006 Pagani Zonda, 268–269
- Paramics, 245–247
- Partial differential equation (PDE), 83, 102
- Photoelectric effect, 289
- Picoscopic modeling
- applications
- connected vehicle technology, 258, 258 f
 - emergency management, 259
 - interactive highway safety design, 257–258
 - transportation forensics, 258–259
- driver, vehicle and environment
- closed-loop system, 253, 254 f
 - driver modeling approach, 253–254, 255 f
 - environment information, 254–255
 - field theory, 255
 - intelligent agent, 254–255
 - interaction, 253
 - roadway traffic, 256
 - speed, acceleration and yaw velocity, 256–257
 - traffic operation, 256–257
 - transportation system, 253–254, 254 f
 - vehicle dynamic responses and lateral movement, 255–256
 - vehicle feedback, 254–255
- Pipes model
- applications of
 - automatic driving, 184–185
 - computer simulation, 185–186
 - California Vehicle Code, 183
 - fundamental diagram, 192 f
 - macroscopic benchmarking, 191–193
 - microscopic benchmarking, 188–191, 190 f , 191 t
 - properties of, 186–187
 - steady-state model, 331–332
- Pipes-Munjal model, 203, 205
- Pneumatic tubes
- advantages, 8
 - data collection, 7, 8
 - disadvantages, 8
 - installation of, 7 f
 - short-term traffic engineering studies, 7
- Poisson arrival pattern, 330–331
- Prigogine's model, 361–362
- Probabilistic models, 360–361
- Psychophysical model
- driver's operational control, 237
 - driving regimes, 239

Psychophysical model (*Continued*)

- illustration of, 238f
- operating point, 239
- psychological activities and physical behavior, 239
- vehicles collision, 237

R

Radio-frequency identification (RFID)

- advantages, 12
 - data collection, 12
 - disadvantages, 13
 - electronic toll collection system, 12, 12f
 - identification and tracking, 12
 - transponders, 12
- Route-choice models, 173–175, 245
- Rule-based model. *see* HUTSIM
- car-following models, 240–241
 - human intelligence, 241
 - illustration of, 241f

S

Simplified theory of kinematic waves

- backward wave propagation, 153–154
- computational algorithm
 - actual departure, 157, 158
 - capacity constraint, 157
 - departure, 157
 - lattice point, 158
 - time and space, 156–157
 - traffic flow, 156–157
 - upstream arrival, 157
- D/D/1 queuing theory, 151
- Derivation V, 151
- forward wave propagation, 152–153
- local capacity, 154
- minimum principle, 155, 155f
- on-ramp and off-ramp queuing models, 159
- segments, 159
- single bottleneck, 155–156
- traffic dynamics, 151
- triangular flow-density relationship, 152

Single-regime models

- equilibrium traffic flow models
 - comparison of, 61f

equilibrium/steady-state relationship, 51–52

flow, speed and density, 52–55

free-flow speed, 58–59

Georgia NaviGAtor, 51

Greenshields model, 55–58

one-equation models, 59

point sensor, 51, 53f

traffic speed, 51–52

video images, 51

IDM, 227–230

Newell nonlinear model, 221–224

Newell simplified model, 226–227

Van Aerde model, 230–233

Speeds, flow-density relationship

free-flow, 122

kinematic wave, 122

shock wave, 122

traffic, 122

T

Three-regime model, 59–62

Traffic flow characteristics

desired traffic

density, point sensor data, 32–33

field data collection, 31

HCM, 31

loop detector/video camera, 31

space mean speed, point sensor data, 32

generalized definition

flow, mean speed and density, 42

HCM, 37

infinitesimal distance, 38, 38f

infinitesimal duration, 39, 40f

point sensor data, 37

space sensor data, 37

time-space rectangle, 40

mobile sensor data

GPS data, 19, 20f

vehicle's speed, 19

vehicle's trajectory, 19–21, 21f

vehicle's travel time, 19–21

point sensor data

electromagnetism, 23–24

headway, 23, 24

loop detector data, 23–24, 23f

- occupancy, 24
 - on time, 24
 - time mean speed, 24
 - time-space diagram, 22, 22*f*
 - traffic count, 22, 24
 - vehicle speed, 24
 - relationships
 - density and spacing, 28
 - flow and headway, 28
 - flow, speed and density, 27
 - occupancy and density, 30–31
 - time mean speed and space mean speed, 29–30
 - space sensor data
 - density, 25–26
 - space mean speed, 26
 - spacing, 25
 - vehicle speed, 26
 - three-dimensional representation
 - density and space mean speed, 48
 - examples, 44*f*
 - intelligent transportation system, 48
 - N-t* diagram, 46*f*
 - N-t-x* region, 47*f*
 - N-x* diagram, 45*f*
 - vehicle trajectories, 46
 - time-space diagram and characteristics, 26, 27*f*, 27*t*
 - Traffic flow, physical basis of
 - electromagnetic phenomena
 - car following, 287–288
 - shying-away effect, 288
 - tailgating, 288
 - Maxwell-Boltzmann distribution, 290*f*
 - mechanics phenomena
 - directional flow, 285–286
 - free flow, 286
 - Newton's third law of motion, 285
 - red light, 286
 - road barriers, 286–287, 287*f*
 - physical laws and social rules, 284
 - statistical mechanics phenomena, 289–291
 - three systems, 285*f*
 - transportation system, 284
 - wave phenomena
 - harmonic wave, 288
 - signal propagation, 289
 - wave-particle duality, 289
 - Traffic sensing technologies
 - classification
 - intrusive, 13
 - mobile sensor, 13
 - nonintrusive, 13
 - off-roadway, 13
 - point sensor, 13
 - space sensor, 13
 - data sources
 - GA400 data, 13–15
 - NGSIM data, 15–17, 16*f*
 - ITS, 3
 - types
 - acoustic/ultrasonic sensors, 9–10, 10*f*
 - aerial/satellite imaging, 10–11
 - GPS, 8–9, 8*f*
 - inductive-loop detector, 3–6
 - pneumatic tubes, 7–8
 - RFID, 12–13
 - VIPS, 6–7, 6*f*
 - Traffic stream model, 314
 - TRANSIMS, 361–362
 - Two-regime linear model, 59–62
- ## U
- Underwood model, 203, 205, 316–320, 332
 - Unified diagram
 - connections, 351–354
 - field theory, 339–340
 - macroscopic equilibrium models
 - Del Castillo and Benítez model, 350–351
 - Drew model, 349
 - GM-associated models, 351
 - IDM (macroscopic), 348
 - Kosonen model, 347*f*
 - Newell model (macroscopic), 348
 - Pipes-Munjal model, 348–349
 - Van Aerde model (macroscopic), 348
 - Wang et al. model, 349–350
 - microscopic car-following models
 - CARSIM model, 344–345
 - Forbes model, 341
 - Gipps model, 342–343
 - GM models, 341–342

- Unified diagram (*Continued*)
- IDM, 343
 - Newell model, 340–341
 - psychophysical model, 345–346
 - rule-based model, 346–347
 - Van Aerde model, 344
- motivation
- acceleration, 338, 339
 - deceleration, 338
 - driver operational control, 338
 - dynamic car-following model, 338
 - gas and brake pedals, 338
 - human factors, 338
 - Newell nonlinear car-following model, 337
 - Newton's second law of motion, 339
 - physics, 338
 - traffic flow models, 337
- picoscopic models, 351
- traffic flow models, 351, 352*f*
- V**
- Van Aerde car-following model, 331–332
- Van Aerde model
- fundamental diagram, 235*f*
 - macroscopic benchmarking, 231–233, 233*t*
 - microscopic benchmarking, 231, 233*t*, 234*f*
 - Pipes model and Greenshields model, 230
 - properties of, 230–231
- Video detection system (VDS), 13–14
- Video image processing system (VIPS)
- advantages, 7
 - data collection, 7
 - detection zone, 6
 - disadvantages, 7
 - roadway traffic, 6
 - video camera, 6
- VISSIM, 245–247
- W**
- Waves
- characteristic method, 92–94
 - characteristics
 - characteristic lines/characteristics, 90
 - domain of dependence, 89 - range of influence, 89
 - conservation law, 99
 - entropy condition, 109–110
 - general solution to wave equations, 87–89
 - gradient catastrophes
 - break time, 99–101
 - compression wave, 101
 - expansion wave, 101
 - infinite gradient, 99–101
 - solution profile, 101, 101*f*
 - time-space points, 99–101 - mathematical representation
 - homogeneity, 85
 - linearity, 85
 - notation, 84
 - order, 84–85 - PDE, 83
- properties
- characteristics, 94–96
 - solution, 96
- rarefaction waves
- characteristics method, 107
 - characteristics without intersection, 105*f*
 - conservation law problem, 107
 - empty area, characteristic filling, 106*f*
 - Rankine–Hugoniot jump condition, 107–108
 - shock wave solution, 108, 108*f*
 - time development of, 106*f*
- shock waves
- characteristic intersections, 102
 - conservation law, 103
 - kinematic waves, 104
 - piecewise solution, 102*f*
 - Rankine–Hugoniot jump condition, 102, 103–104
 - shock path, 102, 104, 104*f*
- solution, wave equation, 90–92
- surface waves, 84*f*
- traffic waves, 84*f*
- traveling waves, 85–86
- traveling wave solutions, 86
- wave front and pulse, 87
- wave phenomena, 83
- wave terminology, 110–111
- Wiedmann model, 345, 345*f*
- “Wrong-way travel,” 165