Lecture Notes in Traffic Flow Theory

A Unified Perspective

Daiheng Ni

Department of Civil and Environmental Engineering
University of Massachusetts Amherst
MA 01003, USA

February 20, 2013

Preface

For years, I have been thinking about writing an introductory book on traffic flow theory. The main purpose is to help readers like me who are new to this subject and do not have much preparation in mathematics and traffic flow. To serve this purpose, I try to make the contents self-containing and assume minimal requirement on 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 UMass Amherst. Hence, the chapters are more like lectures with focused topics, each of which fits in a class meeting. The book consists of four parts.

Part I focuses on traffic flow characteristics. It starts with Intelligent Transportation Systems (ITS) and traffic sensing technologies to illustrate how to quantify traffic flow and collect such data. This is followed by three chapters with in-dept discussion on traffic flow characteristics, based on which their relationships are developed and a few equilibrium traffic flow models are introduced.

Part II is all 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 a preparation, the domain knowledge of traffic flow is integrated into mathematical modeling, resulting in a first-order quasi-linear partial differential equation problem known as LWR model in traffic flow community. Solutions to the problem are introduced including a graphical technique using method of characteristics and numerical techniques involving a few discretization schemes.

Part III is devoted to traffic flow modeling at the *microscopic* level. The emphasis is drivers' car-following behavior involving operational control in the longitudinal direction. A series of car-following models are introduced with varying modeling philosophies and complexity. In order 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 aiming 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 to the observed).

Part IV extends traffic flow modeling to the *picoscopic* level. A modeling framework called driver-vehicle-environment closed-loop system is introduced to capture ultra-fine 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 make control decision on motion in longitudinal and lateral directions. The vehicle model executes its driver's control decision and move 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. Using the field theory as the basis, a unified perspective can be casted on traffic flow theory. 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, a benchmarking effort is made to cross-compare the performance of some of the macroscopic models and microscopic models in the diagram. Meanwhile, a multi-scale modeling approach is presented which involves traffic flow modeling at a spectrum of four levels of detail, namely macroscopic, mesoscopic, microscopic, and picoscopic. The emphasis of multi-scale is to ensure modeling consistency, i.e., 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 can be used by transportation engineering graduate students as an entry-level textbook of traffic flow theory. In addition, civil engineering juniors and seniors may find some in-dept information about traffic flow fundamentals in this book. Further more, applied math 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 and the following are a few of them:

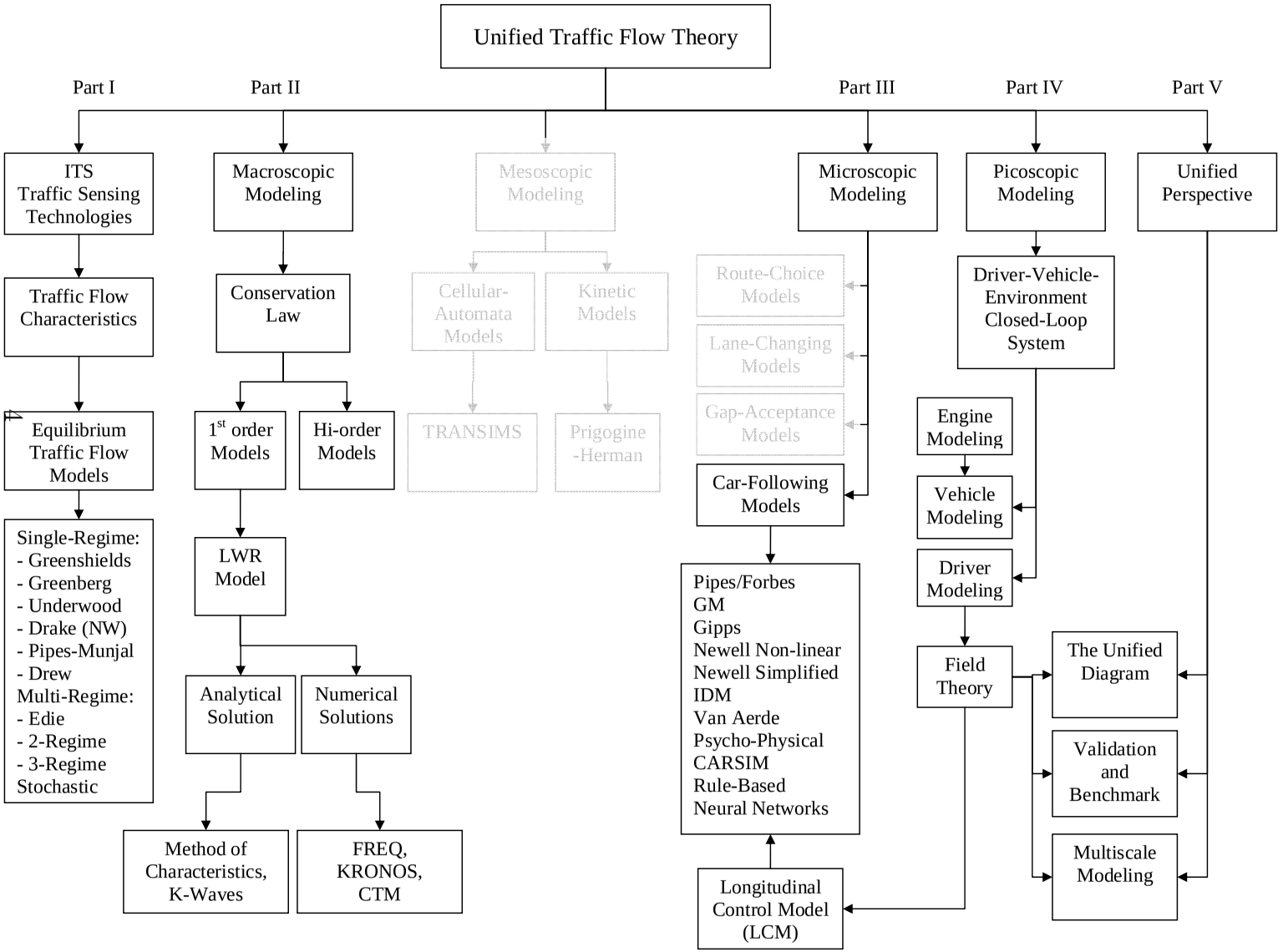
- Adolf D. May. *Traffic Flow Fundamentals*. Prentice-Hall, 1989.

- CF Daganzo. *Fundamentals of Transportation and Traffic Operations*, Pergamon-Elsevier, 1997.
- Gartner et al. *Revised Monograph on Traffic Flow Theory*. TRB 2001. http://www.tft.pdx.edu/docs.htm
- Gerlough and Huber. *Traffic Flow Theory - A Monograph*. TRB 1975. http://www.tft.pdx.edu/docs.htm
- Gerlough and Capelle. *An Introduction to Traffic Flow Theory*. TRB 1964. http://www.tft.pdx.edu/docs.htm
- D.R. Drew. *Traffic Flow Theory and Control*. McGraw Hill, Inc. 1968.
- W. Leutzbach. *Introduction to the Theory of Traffic Flow*. Springer-Verlag, 1988.
- G.F. Newell. *Theory of Highway Traffic Flow*, 1945-1965. Course Notes UCB-ITS-CN-95-1. 1996.

I would like to 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 sparkled 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 should acknowledge that this book is constantly under revision. Though I try hard to ensure the quality of the information, I may make mistakes and, hence, the reader should use this book with discretion.

Daiheng Ni
Amherst, MA
September, 2012



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

Contents

Part I

Traffic Flow Characteristics

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 assistances. Figure 1.1 illustrates a snapshot of a transportation system with various applications of these technologies.

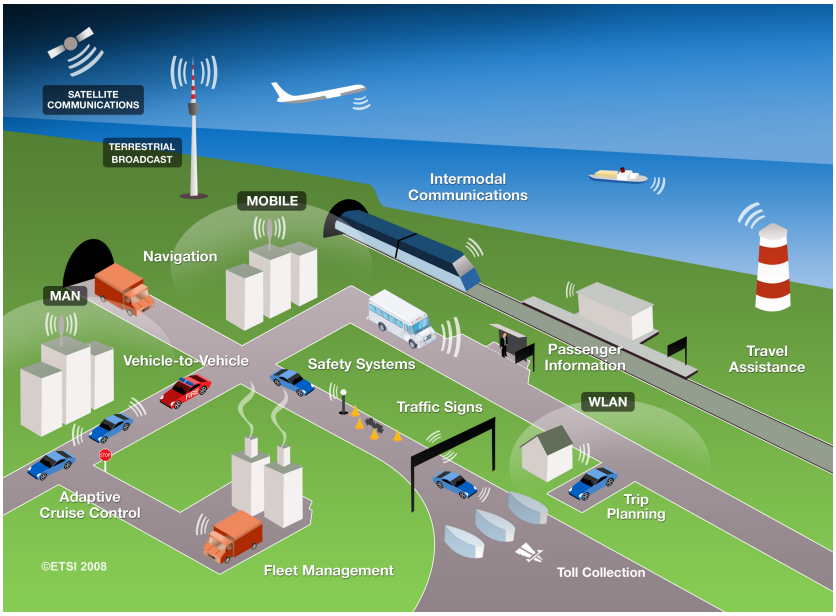


Figure 1.1: An Illustration of ITS (photo credit: www.etsi.org)

1.1 Intelligent Transportation Systems

Intelligent Transportation Systems (ITS) refer 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 applications that address the above objectives: Advanced Traffic Management System (ATMS), Advanced Traveler Information System (ATIS), Advanced Public Transportation System (APTS), Intelligent Vehicle Initiative (IVI), Commercial Vehicle Operation (CVO), etc. Recent development of ITS emphasizes the application of Dedicated Short Range Communications (DSRC) in vehicle-to-vehicle and vehicle-to-roadside wireless communications, i.e., the Connected Vehicle Technology according to the United States Department of Transportation.

The following subsections highlight two sources of data that will be used throughout this book: NaviGAtor and NGSIM.

1.1.1 NaviGAtor - Georgia's ITS

The NaviGAtor is 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 1,645 VDS stations are installed approximately every 1/3 mile along most major Interstates in Atlanta Metropolitan area. These VDS cameras provide continuous speed and volume data to the Traffic Management Center (TMC) and allow the system to generate travel times for the Changeable Message Signs (CMS). NaviGAtor also uses about 500 full-color Closed-Circuit Television (CCTV) cameras, positioned about every 1 mile on most major Interstates in Atlanta. The CCTVs have tilt, pan and zoom capabilities and serve as traffic cameras sending real-time footage to the operators at the TMC for enhanced situational awareness. The information collected from these cameras allows them to confirm incident details, dispatch HERO units, and request appropriate emergency resources.

Figure 1.2 shows a real-time traffic map of NaviGAtor in Atlanta Metropolitan area. On this map, roadway links are color-coded to highlight the level of congestion. In addition, locations of some of the video cameras and change-

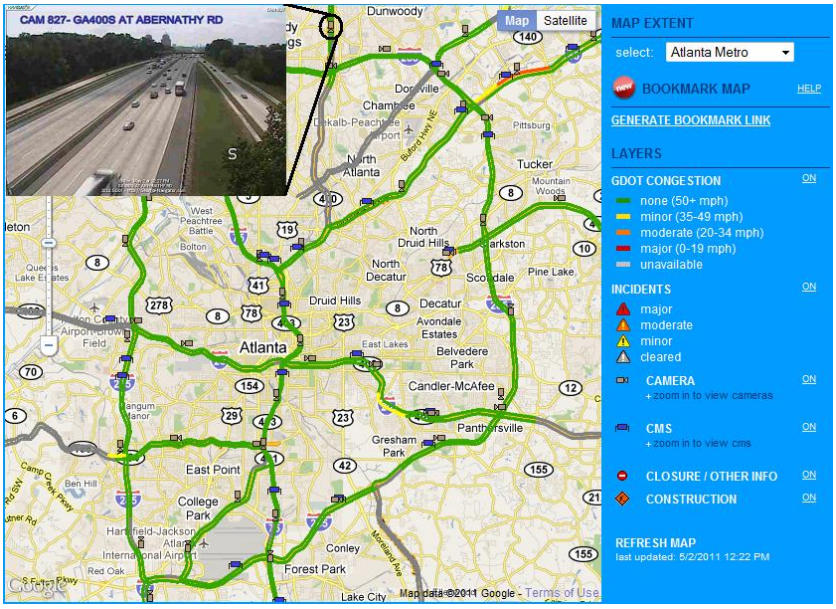


Figure 1.2: Georgia Navigator (photo credit: www.georgia-navigator.com)

able message signs are labeled on the map. A sample image from a video camera on Georgia State Route 400 (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 entry of data represents 20 seconds aggregation of classified vehicle counts, time mean speed, occupancy, etc. Figure 1.3 illustrates three-dimensional traffic density (converted from field data collected on Friday, Oct. 11, 2002) over time and space.

1.1.2 Next Generation Simulation

The Next Generation Simulation (NGSIM) program was initiated by the United States Department of Transportation (US DOT) Federal Highway Administration (FHWA) in the early 2000's. The program developed a core of open behavioral algorithms in support of traffic simulation with a primary focus on microscopic modeling, and collected high-quality primary traffic and trajectory data intended to support the research and testing of the new

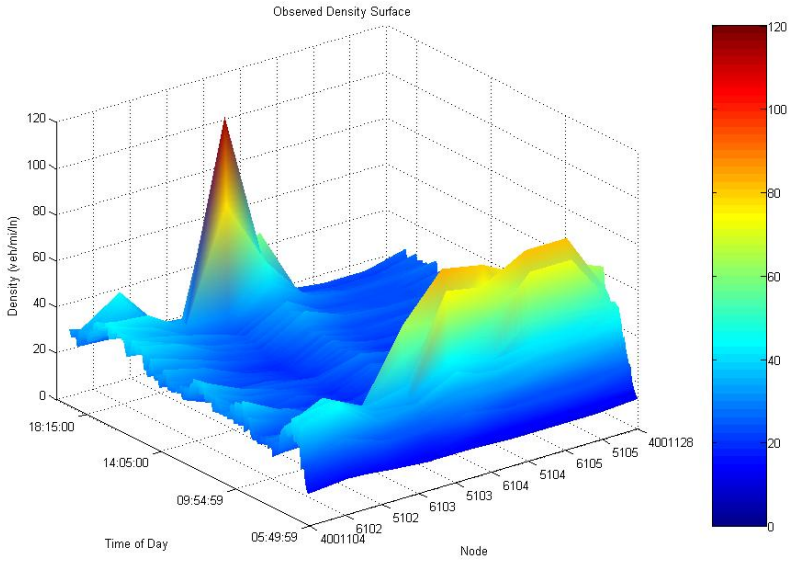


Figure 1.3: GA400 data sample

algorithms. The NGSIM program has actively engaged traffic simulation vendors to accelerate the inclusion of advanced or improved algorithms in the commercial models used across the world.

One of the NGSIM's efforts was to collect detailed vehicle trajectory data on a set of sites including freeways, arterial, and urban streets. Figure 1.4 illustrates one of the sites on I-80 in California. The left pane 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 pane visualizes a camera and its perspective. These cameras shot the site at different angles such that a vehicle entering from the upstream is monitored continuously and consecutively by these cameras till it exits the study area.

Videos captured by these cameras were then processed by a customized software application which was able to identify, track, and record while a vehicle traversed the study areas. The resultant vehicle trajectory data provided the precise location of each vehicle within the study areas every one-tenth of a second, resulting in detailed lane positions and locations relative to other vehicles. Figure 1.5 illustrates a sample result of such vehicle trajectory data. The y-axis (not shown) is the highway running from south to north and the

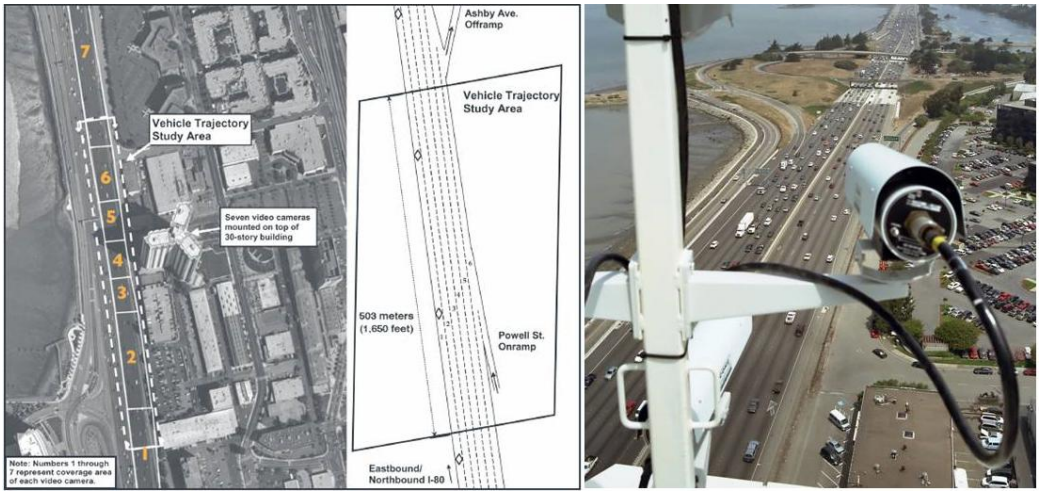


Figure 1.4: NGSIM data collection site (photo credit: www.fhwa.dot.gov)

Figure 1.5: NGSIM data sample

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 as ripples in water.

1.2 Traffic Sensors

This section lists a few types of traffic sensors that are often employed in traffic surveillance and data collection. 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.

1.2.1 Inductive-Loop Detectors

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

How it works

As illustrated in Figure 1.6, an inductive-loop detection system consists of an inductive loop, which is simply a coil of wire embedded in the road's surface, and a detector, which typically sits in a signal cabinet and links the signal controller to the inductive loop. When a vehicle enters or crosses the loop, the body and frame provide a conductive path for the magnetic field. 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 (i.e., an "ON" status). Otherwise, the detector does not output a signal (i.e., an "OFF" status).

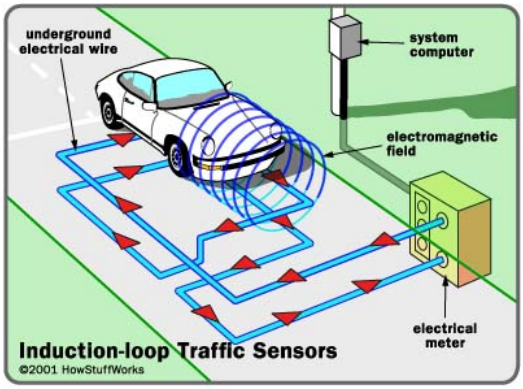


Figure 1.6: An inductive-loop detection system (photo credit: www.ustraffic.net)

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. For another example, when a vehicle exits a gated parking garage, an inductive loop is able to detect the vehicle in advance so that the gate automatically opens for the vehicle. Yet another innovative application is 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 red light, the camera will be triggered and a picture of vehicle will be taken as the evidence of red light violation.

Data collected

An inductive-loop detector monitors a point of roadway and is able to collect (classified) traffic counts, vehicle instantaneous speed, headway (temporal separation between two consecutive vehicles), ON time (time during which the detector outputs an "ON" status), 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 pavement). In addition, set-up and maintenance costs of inductive-loop detectors is high. Inductive-loop detectors can fail under weather conditions, especially snow and ice.

1.2.2 Video Image Processing System

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

How it works

A video image processing system comprises: (1) an image processing system, e.g., a video camera mounted overhead above the roadway that captures real-time images/video streams of the traffic under surveillance, (2) a telecommunication system, e.g., 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 pane of Figure 1.7 illustrates a video camera which is monitoring traffic. The right pane shows an image of roadway traffic (not necessarily a match to the view of the video camera in the left pane) with detection zones set up on the screen. When a vehicle enters a detection zone, the VIPS outputs an "ON" signal which remains till the vehicle exits the detection zone, at which time the VIPS switches to an "OFF" signal. Multiple detection

Figure 1.7: Video image processing system (photo credit: autoscope.com)

zones can be set up, e.g. one for each lane. Hence, these detection zones constitute a detection station.

Data collected

Similar to inductive-loop detectors, the VIPS monitors a point of roadway and is able to collect (classified) traffic counts, vehicle instantaneous speed, headway, ON time.

Advantages

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

Disadvantages

The VIPS is expensive - setup cost is high. It is vulnerable to visual obstruction e.g. inclement weather; shadows, poor-lighting conditions, and strong winds.

1.2.3 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 the 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 1.8: Installation of pneumatic tubes (photo credit: www.arlingtonva.us)

Figure 1.8 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.

Data collected

Pneumatic tubes are able to collect traffic data such as instantaneous speed, direction of flow, volume, vehicle classification, and the time of day associated with each data sample.

Advantages

Pneumatic tubes are portable devices for automatic traffic data collection. The cost is moderate and the system can be reused at other locations. The system requires installation which does not require much efforts (manageable 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.

1.2.4 Global Positioning System (GPS)

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 functionalities, and hence they are considered in the same category of GPS.

How it works

The GPS is a satellite-based navigation system made up of a network of 24 satellites placed into orbit by the U.S. Department of Defense. GPS satellites circle the earth twice a day in a very precise orbit and transmit signal information to the earth. GPS receivers take this information and use triangulation to calculate the user's exact location, see Figure 1.9 for an illustration. Essentially, the GPS receiver compares the time a signal was transmitted by a satellite with the time it was received. The 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.

Figure 1.9: The Global Positioning System (photo credit: wikipedia.org)

Suppose 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.

Advantages

GPS now becomes an affordable technology since one only need a GPS receiver to receive positioning signals. It is simple to install and operate. It works under all weather and lighting conditions.

Disadvantages

GPS receivers only provide 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.

1.2.5 Acoustic/Ultrasonic Sensors

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

How it works

The sensors 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 relative distances between a pedestrian/vehicle and the sensor. In mechanical applications, these sensors can be used to measure fluid levels. The picture here has them installed in the rear of a vehicle as a parking sensor. The sensors measure the distance between the

vehicle and an object behind, and then display a color corresponding to the distance on the dashboard panel. When the display turns red, the driver can stop and be perfectly parked.

Figure 1.10: Acoustic/Ultrasonic Sensors (photo credit: www.liveozshop.com)

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 involve 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 of varying surfaces differently, which may throw off readings on the sensor.

1.2.6 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 1.11, the helicopter can be used to capture images of the ground and the images are stored or transmitted to a work station for analysis. The outcome includes vehicle counts, vehicle speeds, and traffic density.

Figure 1.11: Unmanned helicopter as traffic sensor (photo credit: www.draganfly.com)

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 further mean traffic speed.

Advantages

Traffic surveillance can be taken at high accuracy. There is no need for hardware installation or near roadways, i.e., it is a non-intrusive and non-interruptive technology. It can provide bird's eye view of system-wide traffic conditions.

Disadvantages

Helicopters are expensive and require pilots to operate. It is time- and resource-consuming to collect traffic data. Analysis of aerial photos is complicated, e.g. aligning aerial photos captured from different angles.

1.2.7 RFID Technology

Radio-frequency identification (RFID) is the core technology of many traffic sensors know as transponders (e.g., E-ZPass tags), Automatic Vehicle Identification (AVI), etc.

How it works

Radio-frequency identification (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 1.12 illustrates an electronic toll collection (ETC) 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 at a speed, the tag reader detects the transponder and records the its unique ID, the time instant, and other account-related information such as balance and toll paid.

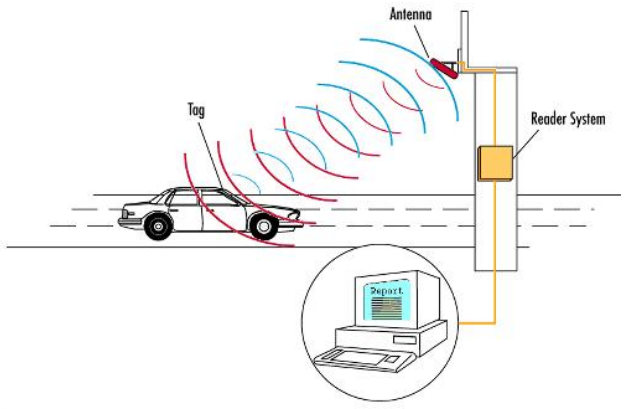


Figure 1.12: Electronic toll collection system (photo credit: HowStuffWorks)

Data collected

RFID is able to record and the IDs of equipped vehicles and time-stamp the arrival of such vehicles.

Advantages

RFID is inexpensive. It does not interrupt traffic.

Disadvantages

RFID only detects equipped vehicles at a point of roadway.

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

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

- *intrusive* if installation of the sensing system requires pavement work and interrupting traffic. Examples of intrusive sensors are inductive-loop detectors and pneumatic tubes.

- *non-intrusive* if installation of the sensing system does not require pavement work and interrupting traffic. Examples of intrusive sensors are video image processing systems and RFID technology.
- *off-roadway* if the sensor is not fixed to a location of the roadway, i.e. the sensor can move with vehicles or float in the sky. Examples of space sensors are GPS receivers, acoustic/ultrasonic sensors, cell phones, helicopters, satellites.

Chapter 2

Traffic Flow Characteristics I

We have learned that, 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/on a vehicle, moves along with the vehicle, and reports the locaion of this particular vehicle over time. A point sensor sits at a fixed location along or above a roadway, sees the passage of vehicles in front of it, and reports traffic data only at this particular location over time. A space sensor flies in the sky, sees a stretch of road underneath, and records positions of vehicles on this particular stretch of road at an instant of time.

2.1 Characteristics and mobile sensors

Next, let's take a look at how traffic data reported by these sensors look like. First, mobile sensors. If a vehicle is equipped with a GPS device, the device can report the vehicle's position as time moves on. Since GPS typically reports once every second, the GPS data may look similar to Table 2.1 where the vehicle's longitudinal x and lateral y displacements are in relative to the vehicle's position at 9:00:00.

Figure 2.1 shows the scenario in which a vehicle (with an on-board GPS device and is numbered i) is moving on a roadway (in the left) and the associated time-space diagram (in the right). Every circle represents represents a GPS reading (only x is shown and y ignored). If one connects these circles, one obtains the **trajectory** of this vehicle, i.e. 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,

Table 2.1: GPS data

Time	X (ft)	Y (ft)
9:00:00	0	0
9:00:01	3	0
9:00:02	5	0
9:00:03	7	0
9:00:04	10	1
9:00:05	15	4
9:00:06	18	9
9:00:07	21	12
9:00:08	23	12
9:00:09	27	12
9:00:10	30	12

$\dot{x}_i$, from the GPS data as illustrated in the figure.

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

If the vehicle trajectory is known and smooth, $\dot{x}_i$ can be determined by taking the first derivative of the trajectory:

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

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

$$t_i = t_i^B - t_i^A$$

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. Subplot (a) is valid and it shows a vehicle moving in the positive x direction over time. Subplot (b) is not a valid trajectory because of the following. 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

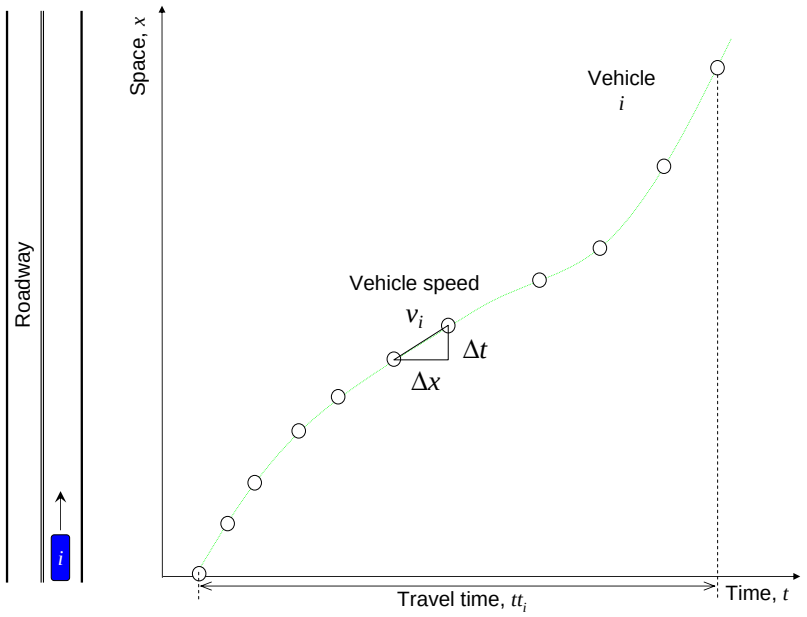


Figure 2.1: A trajectory

the same reason, (c) and (j) are invalid either. (d) is a valid one and the trajectory suggests that the vehicle fist moves forward (i.e. in positive x direction) and then, at some point of time, the vehicle reverses. (e) is valid which simply suggests that the vehicle does not move (maybe parked). (f) is impossible because it suggests an infinite speed (i.e. the tangent of the trajectory). (g) is a valid one since the vehicle just moves backwards at a time-varying speed. (h) is likely but very unusual because the vehicle first moves at reasonable speeds and then almost flies at the end. (i) is valid and the vehicle gradually comes to a stop. (k) can be interpreted in two ways. One is a two-lane scenario where a fast vehicle overtakes a slow vehicle. Two is a one-lane scenario where the fast vehicle crashes with the slow vehicle and they exchange momentum. (l) suggests that two vehicles crash into one and move as a single unit thereafter.

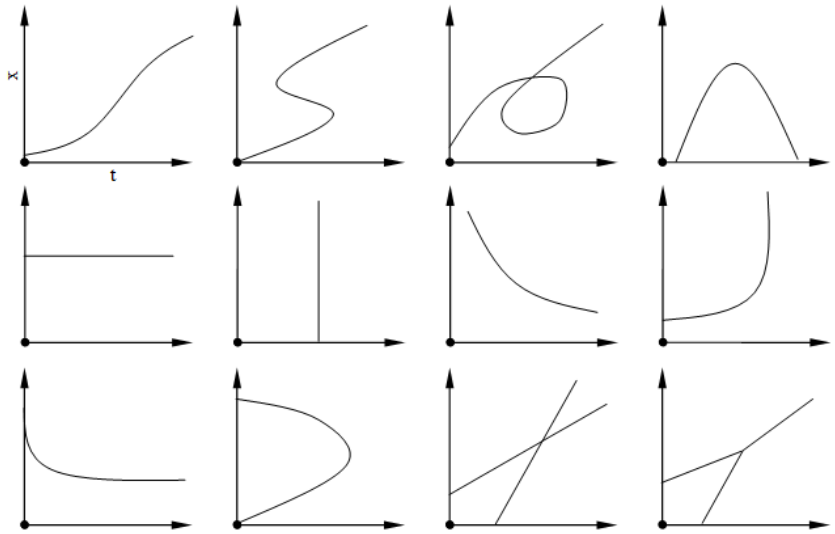


Figure 2.2: Vehicle trajectories

2.2 Characteristics and point sensors

If a point sensor (such as a loop detector or a video camera) is placed on the road at location x, this sensor will be able to observe vehicles passing through it. In a time-space diagram as illustrated in Figure 2.3, each vehicle will be counted (eg. 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 **traffic count** which can be converted to the rate of traffic flow (referred as "**flow**" q thereafter) as follows:

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

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:

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

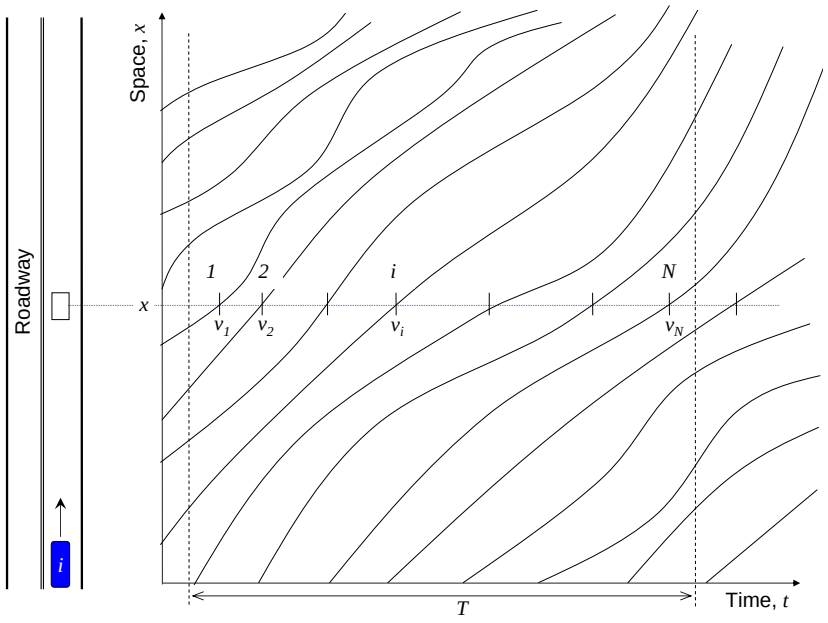


Figure 2.3: Point sensor data

It is true that both a vehicle and a point sensor have physical dimensions. If the sizes of 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 hits a loop, eddy current will be generated in the loop according to electromagnetism. When the vehicle's rear bumper exits the loop, the current drops. See an illustration of this effect in the lower plot above the two trajectories in Figure 2.4. If a threshold is properly set, the loop detector outputs two states: ON when a vehicle is above the loop and OFF then the loop sees no vehicle over it. See an illustration in the upper plot above the two trajectories in Figure 2.4. 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:

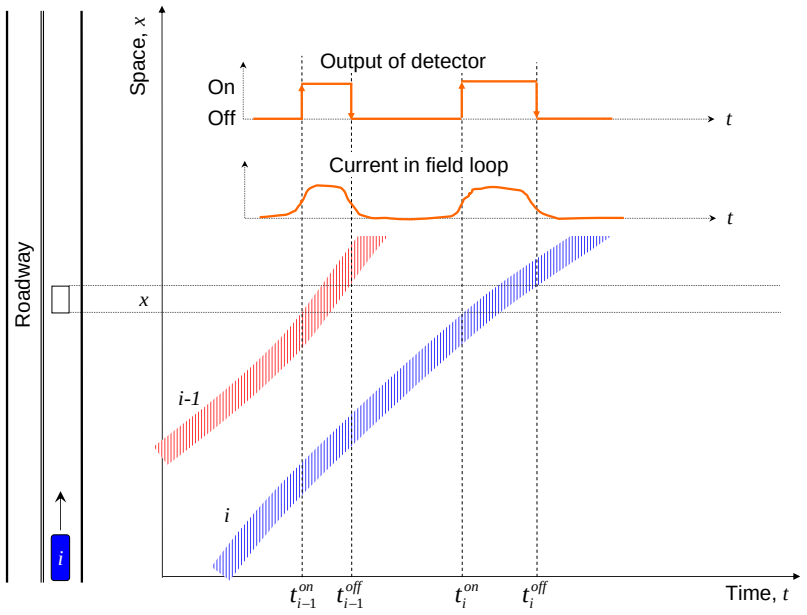


Figure 2.4: Loop detector data

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

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

ON time τ_i: the duration from the moment when a vehicle's (eg. i) front bumper hits the loop to the moment when the vehicle's rear bumper exits the loop: $\tau_i = t_i^{off} - t_i^{on}$.

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 for small loops) and l_i is the length of the vehicle. Hence, the vehicle's instantaneous speed can be determined as

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

Occupancy o: In traffic flow theory, the term occupancy is defined as

the percent of time when a loop is busy, ie. the loop detects vehicles above it. Hence, if the observation period is T, during which N vehicles are detected, the total ON time is $\sum_{i=1}^{N} \tau_i$ and occupancy is determined as

$$o = \frac{\sum_{i=1}^{N} \tau_i}{T}$$

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 the time-mean speed.

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

Interested readers are referred to reference [116] where detailed discussion is given on how various traffic flow characteristics are measured and calculated as well as how errors inherent in point sensors are introduced.

Chapter 3

Traffic flow characteristics II

3.1 Characteristics and space sensors

If one takes aerial photos of a roadway from a helicopter, one is able to locate vehicles in each of these snapshots. Figure 3.1 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 error due to incomplete spacings of the first and last vehicles, the length of roadway L can be expressed as:

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

One is unable to determine **Vehicle speed** v_i from a single snapshot, but with two snapshots, one is able to compare vehicle locations and find the

distance traversed by each vehicle. Since the time between the two snapshots is known, the speed of each vehicle can be determined accordingly.

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

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

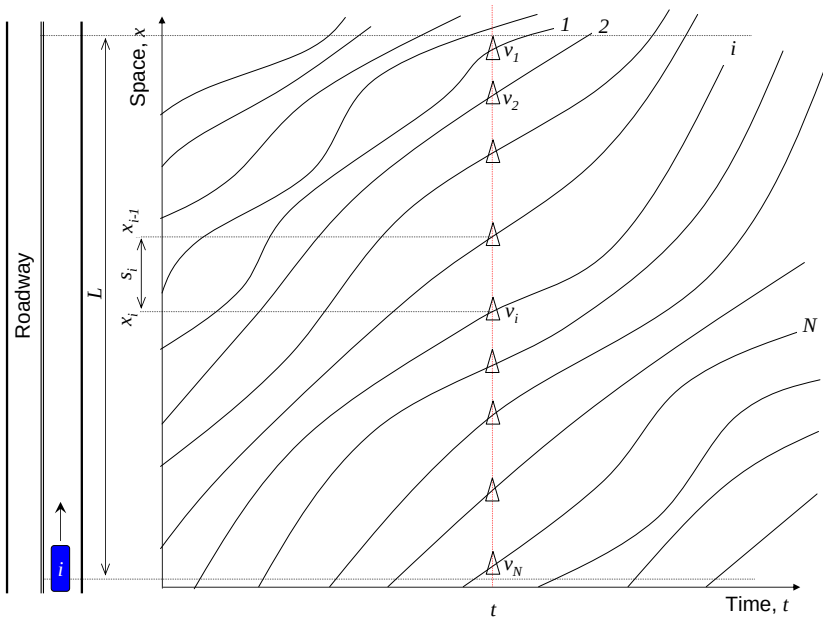


Figure 3.1: A snapshot of roadway

3.2 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 using these data. It is informative to put everything together and form a complete picture. Figure 3.2 shows a time-space diagram with vehicle trajectories where data reported by the three types of sensors are illustrated.

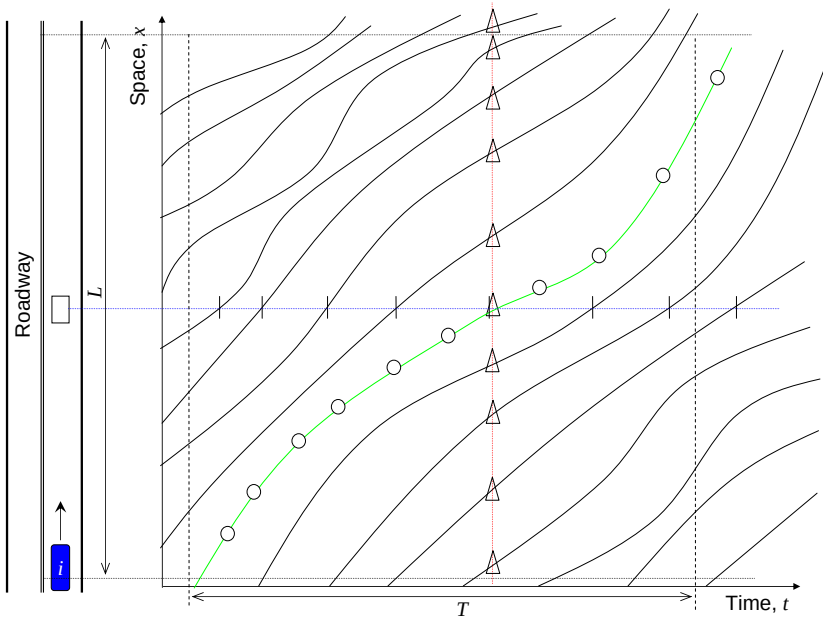


Figure 3.2: Time-space diagram and 3 types of sensors

Table 3.1 relates traffic flow characteristics to sensor types. Three categories of traffic flow characteristics are presented: flux, speed, and concentration. 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 taken over vehicle, time, or space.

3.3 Relationships among characteristics

3.3.1 Flow, speed, and density

By definition, the following relationship holds as an identity:

$$q = k \times v_s$$

i.e., flow q is the product of density k and time mean speed v_s.

Table 3.1: Sensors and traffic flow characteristics

Category	Sensors	Micro Character	Macro Character
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	τ_i	o
	Space	s_i	k

3.3.2 Flow and headway

Based 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 (vph) suggests an average headway of

$$\frac{1}{1200 \text{ vph}} = \frac{3600 \text{ sec/hr}}{1200 \text{ veh/hr}} = 3 \text{ sec}$$

3.3.3 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 (vpm) suggests an average spacing of

$$\frac{1}{40 \text{ vpm}} = \frac{5280 \text{ ft/mi}}{40 \text{ veh/mi}} = 132 \text{ ft}$$

3.3.4 Time-mean speed and space-mean speed

Below is 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 (mph) in the inner lane and slow vehicles all move at 30 mph in the outer lane. traffic flow in each lane is 1200 vph and lane change is prohibited. What is the time-mean speed and space-mean speed?

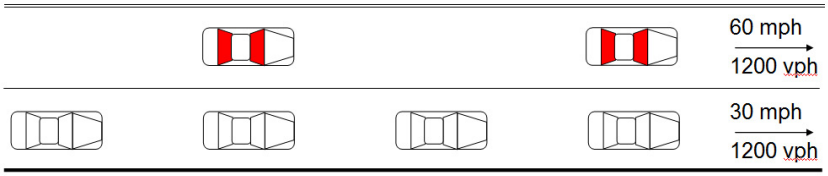


Figure 3.3: Time-mean speed vs. space-mean speed

Calculation of space-mean speed is straightforward, one simply averages speed of vehicles observed on the road. Within one mile of the road, one observes a total of 60 vehicles, of which 20 vehicles are in the inner lane (1200 vph / 60 mph) and 40 vehicles in the outer lane (1200 vph / 30 mph). Therefore, space-mean speed is determined as

$$v_s = \frac{20 \times 60 \text{ mph} + 40 \times 30 \text{ mph}}{60} = 40 \text{ mph}$$

For time-mean speed, one has to imagine a hypothetical observer standing at roadside watching vehicles passing in front of her. As a result, she records

2400 vehicles in one hour, of which 1200 vehicles are in the inner lane and 1200 vehicles in the outer lane. Hence by definition, time-mean speed is

$$v_t = \frac{1200 \times 60 \text{ mph} + 1200 \times 30 \text{ mph}}{2400} = 45 \text{ mph}$$

So the results show that the two means are not equal. Actually, time-mean speed is greater than space-mean speed. Wardrop [154] further demonstrated that the following relationship always holds between time-mean speed and space-mean speed:

$$v_t = v_s + \frac{\sigma^2}{v_s}$$

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

Please note that, in the above example, fast vehicles are over represented in time-mean speed (fast : slow = 1 : 1), while in reality the correct ratio is (fast : slow = 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.

3.3.5 Occupancy and density

The following is reproduced from Chapter 2 of Revised Monograph on Traffic Flow Theory [42]:

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

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

3.4 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. Space sensors do not fit 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, 7 days a week. Let alone the complexity of and time spent extracting traffic data from the huge amount of aerial photos. Therefore, the only type of sensors that is suited for regular, large-scale applications is point sensors such as loop detectors and video cameras, see Chapter 1 for details.

Meanwhile, traffic flow characteristics are not equally attractive in terms of traffic control and management. 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. Highway Capacity Manual (HCM) use density as the measure of effectiveness (MOE) to determine level of service (LOS) on freeways and multilane highways.

Hence, there comes the 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 in large scale on a regular basis, while point sensors are permeant (most automated traffic surveillance systems use point sensors) which report less attractive traffic flow characteristics such as time-mean speed and occupancy. Therefore, a common practice, inevitably though, is to estimate space-based

characteristics from point sensor data.

3.4.1 Determining space-mean speed from point sensor data

If a point sensor system is able to keep track of individual vehicle speeds $(\dot{x}_i, i = 1, 2, ..., N)$, 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 arithmetic mean (in this case the time-mean speed). Unfortunately, many point sensor systems only log 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.

3.4.2 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 timestamps the passage vehicles at two locations of the road with no vehicle appearing or exiting in between (e.g. a tunnel), it is possible to construct a curve showing 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 Reference [106] for further details.

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

- $k = \frac{1}{s}$ if average spacing s is known;
- $k = \frac{q}{v_s}$. For point sensor data, replacing v_s with v_t sometimes yields more accurate k than that estimated from occupancy.
- estimating density from travel times using Kalman Filter technique [142].

Chapter 4

Traffic flow characteristics III

4.1 Generalized definition

In previous chapters, flow q and time-mean speed v_t are defined with the help of Figure 2.3 based on point sensor data.

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

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

Similarly, density k and space-mean speed v_s are defined using Figure 3.1 based on space sensor data.

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

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

However, there is no common ground between the two sets of definition and this becomes more evident in Figure 3.2 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 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 definition. Therefore, one is unable to conclude that the identity

$$q = k \times v_s$$

is guaranteed by definition. Therefore, the key issue here 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 thereafter since the definition is formally given in HCM.

In order to find the common ground, let us rearrange the above definition as follows:

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

where dx denotes an infinitesimal distance, see Figure 4.1. 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 distances traversed by all vehicles during time period T:

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

and 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:

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

By definition, the mean speed of vehicles, v, is the total distance traveled by all vehicles divided by the total travel time spent by 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,

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

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

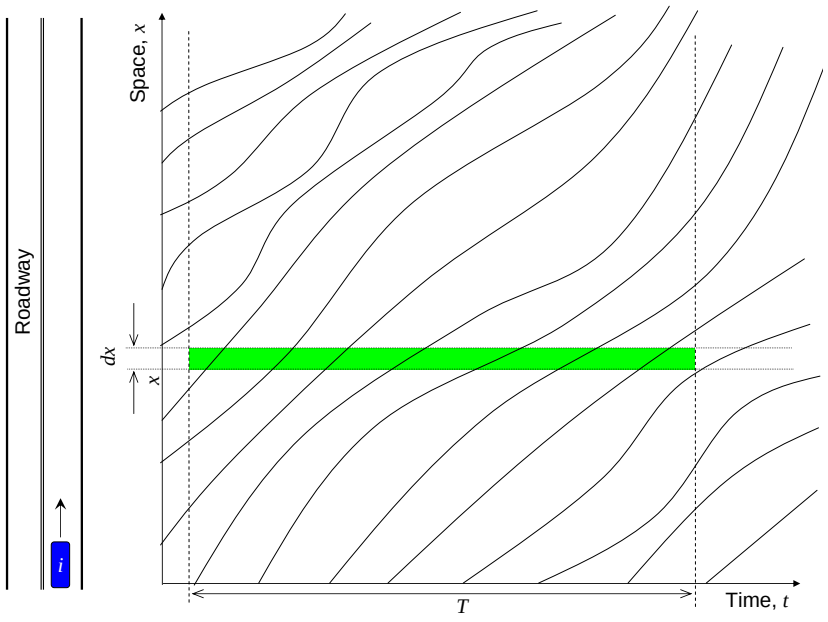


Figure 4.1: Time-space diagram with infinitesimal distance

Similarly, density k can be treated as:

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

where dt denotes an infinitesimal duration, see Figure 4.2. Following the same argument as above, L and dt defines a time-space rectangle A. The numerator is the sum of times spent by all vehicles within A, $t(A)$, and the denominator is the area of the rectangle, $|A|$:

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

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

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

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

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

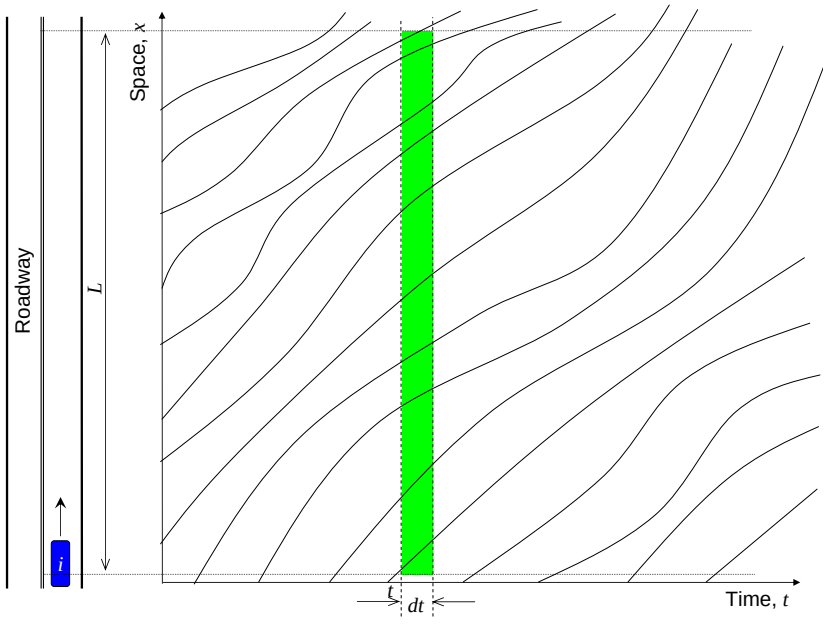


Figure 4.2: Time-space diagram with infinitesimal duration

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 4.3 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}). Based on A, the three traffic flow characteristics can be defined as follows:

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

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

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

Therefore, the identity $q = k \times v$ is now guaranteed by definition. The question 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 late and at time t_{lo} or $t_i(x_{lo})$ whichever comes late; the vehicle exits A at location x_{hi} or $x_i(t_{hi})$ whichever comes early and at time t_{hi} or $t_i(x_{hi})$ whichever comes early. 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 flow, mean speed, and density are determined.

A 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 4.4. The above definition was

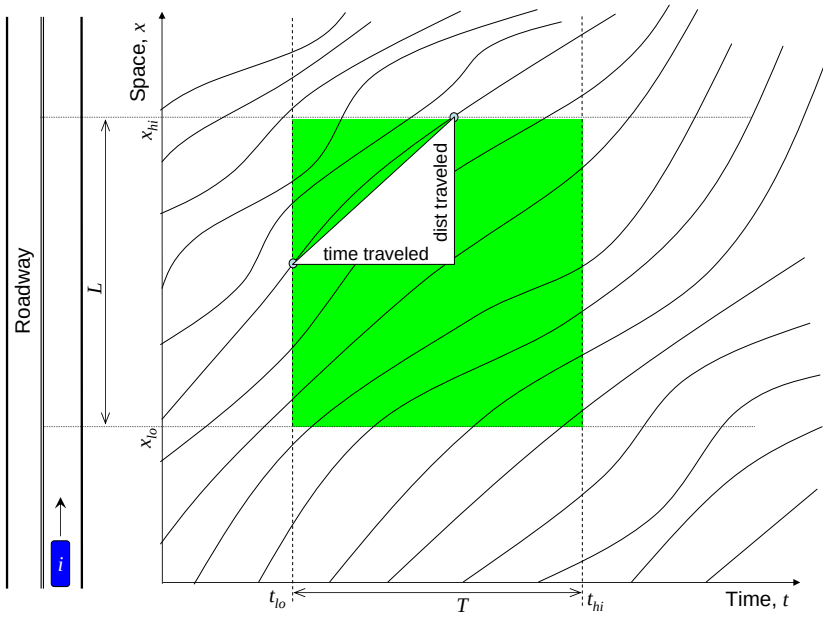


Figure 4.3: Time-space diagram with rectangle

originally proposed by Edie [33]. Readers are referred to the original paper for in-depth discussion. For convenience, the above set of definition of flow, mean, and speed 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 4.1 and let $dx \to 0$, a point sensor scenario is resulted, while obtains the space sensor scenario in Figure 4.2 if one lets $dt \to 0$.

4.2 3D representation

The following discussion is based on reference [86]. Interested readers are referred to the original paper for in-depth information.

So far, we have been working on time-space diagram and vehicle trajectories, based on which connection is drawn to traffic flow characteristics. Time-space diagram is a two-dimensional (2D) representation and the discussion can be made more informative if we take a three-dimensional (3D)

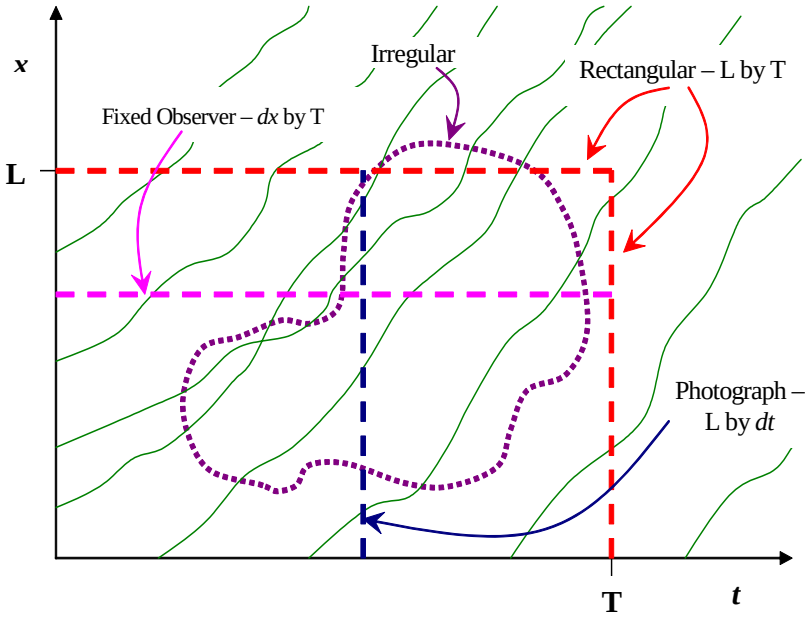


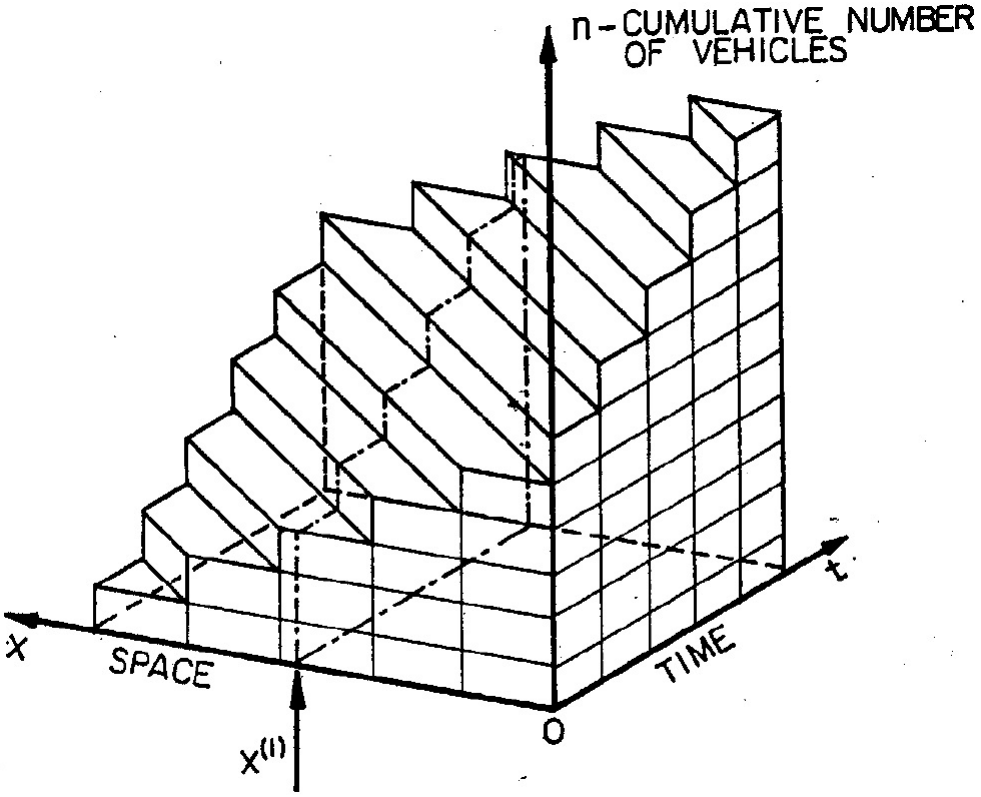
Figure 4.4: Time-space diagram with general region

perspective. Taking the family of vehicle trajectories in Figure 4.3 for example, these trajectories lie on the same plane defined by time (t) an space (x). If 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 number (ie. vehicle 1 raised to height 1, vehicle 2 raised to 2, and so on). Let's call the third dimension the cumulative number of vehicles (N) and denote the surface that passes these elevated vehicle trajectories $N(x, t)$. Figure 4.5 illustrates two examples of such 3D representation adopted from [86, 30] [1] [2].

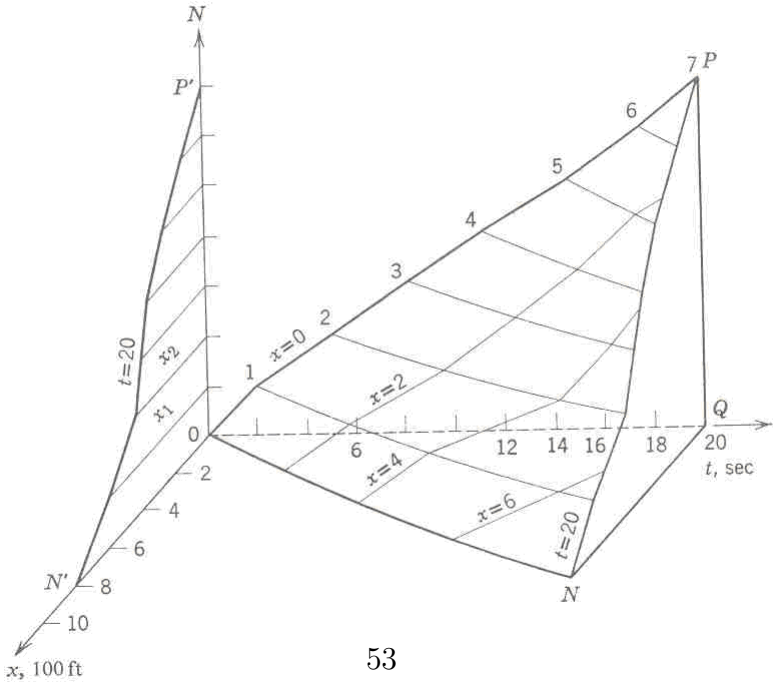
What makes this 3D representation interesting is that it can be used to illustrate and relate some key concepts of traffic flow conveniently. For example, if one cuts $N(x, t)$ in the lower part of Figure 4.5 using plane $t = 20$, one obtains the shape PQN and its projection P'Q'N' on the N-x plane. Curve P'N' can be interpreted as the snapshot taken at time $t = 20$ which

[1] N(x,t) in the upper part of Figure 4.5 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 multi-valued at a time-space point. Makigami [86] showed how to resolve the problem



(a) An example in (Makigami 1971)



(b) Another example in (Drew 1968)

Figure 4.5: 3D representation examples

shows the location of each vehicle at this moment. Figure 4.6 illustrates more examples of such curves (they look like stairs before smoothing) by cutting $N(x,t)$ in the upper part of Figure 4.5 at different instants and projecting 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 runs a horizontal line at height $N = 2$, the intersection of this line and the curve labeled $t^{(1)}$ (actually this line needs to be slightly lower, say at height $N = 1.999$, to avoid multiple intersections, the same thereafter) is the location of vehicle with ID number 2 at time $t^{(1)}$. Similarly, the intersection of line $N = 2$ and curve $t^{(2)}$ is the location of vehicle 2 at time $t^{(1)}$. the length between 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 4.5), 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 bears negative value. Hence,

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

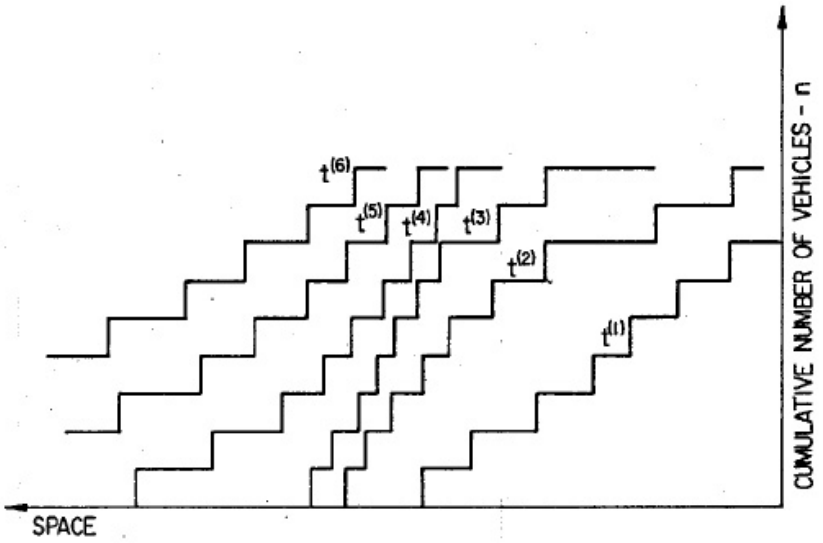


Figure 4.6: The N-x diagram in (source: Makigami 1971)

Similarly, if one cuts the 3D 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 time. For example, the curve in the lower part of Figure 4.5 and curves in Figure 4.7. If one runs a horizontal line at height $N = 2$ in Figure 4.7, the intersection of this line and the curve labeled $x^{(2)}$ indicates the time when vehicle with ID number 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 length between the two intersections is the travel time for vehicle 2 to traverse from location $x^{(2)}$ to $x^{(3)}$. If an N-t curve at location x is smoothed, the tangent of this curve denotes the flow at this location:

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

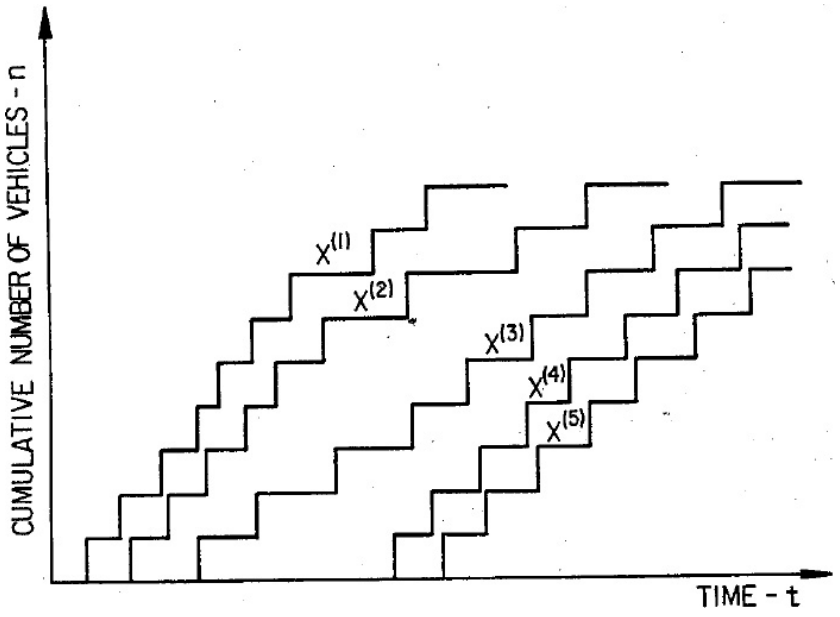


Figure 4.7: The N-t diagram (source: Makigami 1971)

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 $N(x,t)$ (eg. region A in Figure 4.8) onto the x-t, N-t, and N-x planes, one obtains three projections A_N, A_x, and A_t respectively. Makigami [86] demonstrated that the following relationships hold:

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

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

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

$$v = \frac{A_t}{A_x}$$

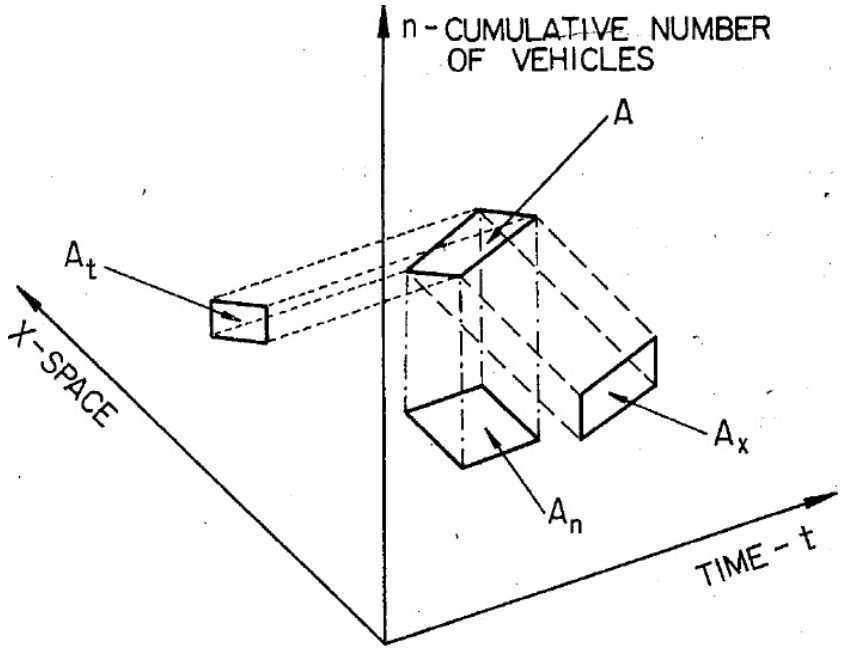


Figure 4.8: Projection of an N-t-x region (source: Makigami 1971)

Figure 4.9 summarizes the above graphics in one figure. Subplot A shows vehicle trajectories in the x-t plane. Subplot D raises vehicle trajectories to their corresponding height and forms the 3D surface $N(x,t)$. Subplot B

shows two N-t curves observed at locations $x = x_2$ and $x = x_4$. Subplot C depicts two N-x curves resulted from snapshots taken at $t = t_6$ and $t = t_8$.

In addition to deepening the understanding of traffic flow and its characteristics, the 3D 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, determining these traffic flow characteristics based on generalized definition (as opposed to HCM definition) is also preferred. However, the widely deployed Intelligent Transportation Systems (ITS) consist mostly of point sensors which are generally unable to report space-based traffic flow characteristics. Interested readers are referred to reference [106] to learn how the 3D representation helps address the problem by computing the desired traffic flow characteristics based on the preferred definition from ITS data.

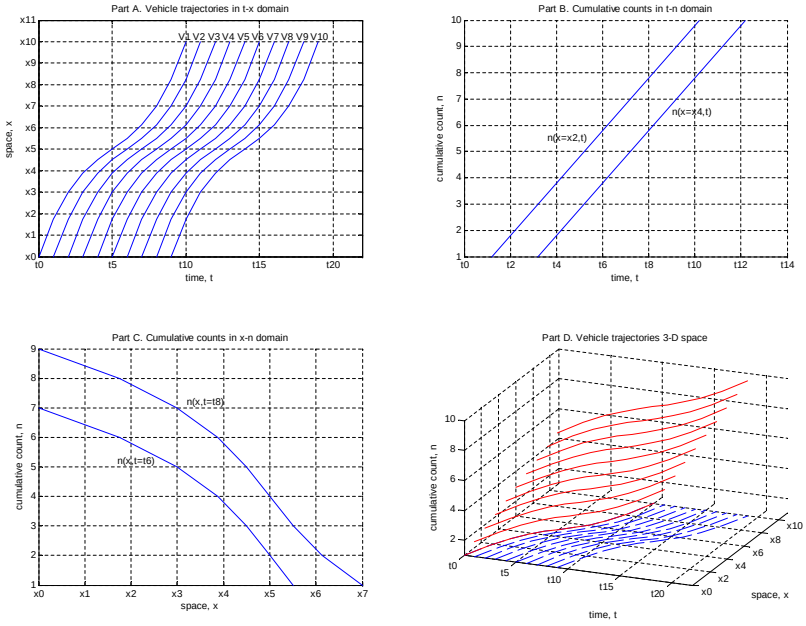


Figure 4.9: 3D representation of traffic flow

Chapter 5

Equilibrium Traffic Flow Models

In previous chapter, it was shown that the following relationship holds among flow q, density k, and space-mean speed v (the subscript s is dropped unless it is necessary to distinguish v_s from time-mean speed v_t):

$$q = k \times v$$

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

5.1 Single-regime models

Let us start with field observations. Figure 5.1 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 5.2 shows a portion of a daily report from a video camera. Each row represents an observation aggregated over 20 seconds over all lanes. Column A is station id; column B is the timestamp of each observation; column C is the status of detectors of this station (there are

Figure 5.1: An image captured by a point sensor (image source: NaviGAtor)

4 lanes at this station and hence 4 detectors. OK means the corresponding detector is working properly while NO_ACT means no actuation); columns E through H are classified traffic counts (only column E is shown here due to limited space); column I is occupancy; column K is time-mean speed; column M is average vehicle length; column P is density (estimated using a proprietary recipe).

The point sensor data are plotted in Figure 5.3 where the top left subplot shows speed-density relationship, the top right speed-flow relationship, the bottom left flow-density relationship, and the bottom right speed-spacing relationship. A few notes about the Figure are elaborated as follows. The the "cloud" represents one year worth of field observations aggregated to five minutes (i.e, each point in the figure represents traffic condition observed in five minutes). The traffic speed here is time-mean speed since it is impossible to calculate space-mean speed from aggregated point sensor data. Density is estimated from flow and speed. The large dots represent the average of the "cloud". Also note that the figure is generated using traffic data collected at a fixed location. Therefore such a plot is location-specific, i.e. plots generated from different locations may vary. In addition, time information is lost in the figure, i.e. one could not deduce the time when a data point was observed.

	A	B	C	E	I	K	M	P	Q
1	#station_id	sample_start	status	volume_auto	time_occupancy	time_speed	length	density	gap
803	4001134	2003-10-08 05:41:00.000 EDT	|NO_ACT|OK|OK|OK	6	0.014	49.1503	14.7638	8.1	796.9
804	4001134	2003-10-08 05:41:20.000 EDT	|OK|OK|OK|OK	23	0.0276	47.4313	14.5451	13.873	383.8095
805	4001134	2003-10-08 05:41:40.000 EDT	|OK|OK|OK|OK	17	0.031	48.2745	14.5607	15.5	353.4524
806	4001134	2003-10-08 05:42:00.000 EDT	|OK|OK|OK|OK	20	0.0473	46.1478	14.7638	18.6429	292.9821
807	4001134	2003-10-08 05:42:20.000 EDT	|OK|OK|OK|NO_ACT	9	0.0136	53.0384	14.7638	6.7143	761.4286
808	4001134	2003-10-08 05:42:40.000 EDT	|OK|OK|OK|OK	14	0.02	48.8921	14.7638	11.1579	895.8947
809	4001134	2003-10-08 05:43:00.000 EDT	|OK|OK|OK|OK	8	0.0086	45.0776	14.7638	5.8182	1055.682
810	4001134	2003-10-08 05:43:20.000 EDT	|OK|OK|OK|OK	13	0.0252	50.8922	13.8642	10.7742	538.0645
811	4001134	2003-10-08 05:43:40.000 EDT	|OK|OK|OK|OK	8	0.0221	52.8492	14.7638	11.2632	619.4211
812	4001134	2003-10-08 05:44:00.000 EDT	|OK|OK|OK|NO_ACT	10	0.0138	54.9913	14.7638	7.375	779
813	4001134	2003-10-08 05:44:20.000 EDT	|OK|OK|OK|OK	17	0.019	50.2191	14.7638	10.7	1558.74
814	4001134	2003-10-08 05:44:40.000 EDT	|OK|OK|OK|OK	14	0.0446	44.9162	14.7638	19.8	295.2571
815	4001134	2003-10-08 05:45:00.000 EDT	|OK|OK|OK|OK	24	0.0557	43.5066	17.0135	18.2857	295.1857
816	4001134	2003-10-08 05:45:20.000 EDT	|OK|OK|OK|OK	15	0.0372	49.8987	14.7638	17.6522	439.5217
817	4001134	2003-10-08 05:45:40.000 EDT	|OK|OK|OK|OK	8	0.0115	50.9524	14.7638	6.7	1667.55
818	4001134	2003-10-08 05:46:00.000 EDT	|OK|OK|OK|OK	18	0.0229	49.7925	14.7638	11.5556	461.1111
819	4001134	2003-10-08 05:46:20.000 EDT	|NO_ACT|OK|OK|OK	8	0.0236	51.4406	14.7638	10.2143	544.3571
820	4001134	2003-10-08 05:46:40.000 EDT	|OK|OK|OK|OK	6	0.0092	46.555	14.7638	4.1538	1429.615
821	4001134	2003-10-08 05:47:00.000 EDT	|OK|OK|OK|OK	14	0.0239	51.5573	14.7638	12.1842	1225.237
822	4001134	2003-10-08 05:47:20.000 EDT	|OK|OK|OK|OK	24	0.04	45.6946	14.7032	21.2769	330.6923
823	4001134	2003-10-08 05:47:40.000 EDT	|OK|OK|OK|OK	11	0.0309	44.9755	19.685	11.7812	484.8125
824	4001134	2003-10-08 05:48:00.000 EDT	|OK|OK|OK|OK	6	0.0094	49.6365	14.7638	4.8824	1318
825	4001134	2003-10-08 05:48:20.000 EDT	|OK|OK|OK|OK	8	0.0125	49.4766	14.7638	4.875	1271
826	4001134	2003-10-08 05:48:40.000 EDT	|OK|OK|OK|OK	16	0.0495	47.9935	14.7638	23.0476	253.7381
827	4001134	2003-10-08 05:49:00.000 EDT	|OK|OK|OK|OK	26	0.0421	47.6679	14.7638	20.1429	578.4935
828	4001134	2003-10-08 05:49:20.000 EDT	|OK|OK|OK|OK	23	0.0577	49.0984	16.0546	20.3115	265.3279
829	4001134	2003-10-08 05:49:40.000 EDT	|OK|OK|OK|OK	20	0.0412	47.9673	14.7638	20.451	268.1961
830	4001134	2003-10-08 05:50:00.000 EDT	|OK|OK|OK|OK	14	0.0334	44.5611	14.3701	13.9714	464.8857
831	4001134	2003-10-08 05:50:20.000 EDT	|OK|OK|OK|OK	15	0.0303	48.8531	14.7638	13.7027	404.027
832	4001134	2003-10-08 05:50:40.000 EDT	|OK|OK|OK|OK	25	0.0488	45.2629	14.7638	24.7344	250.25
833	4001134	2003-10-08 05:51:00.000 EDT	|OK|OK|OK|OK	23	0.0549	44.1544	14.7638	27.6567	199.3881

Figure 5.2: Point sensor data (data source: NaviGAtor)

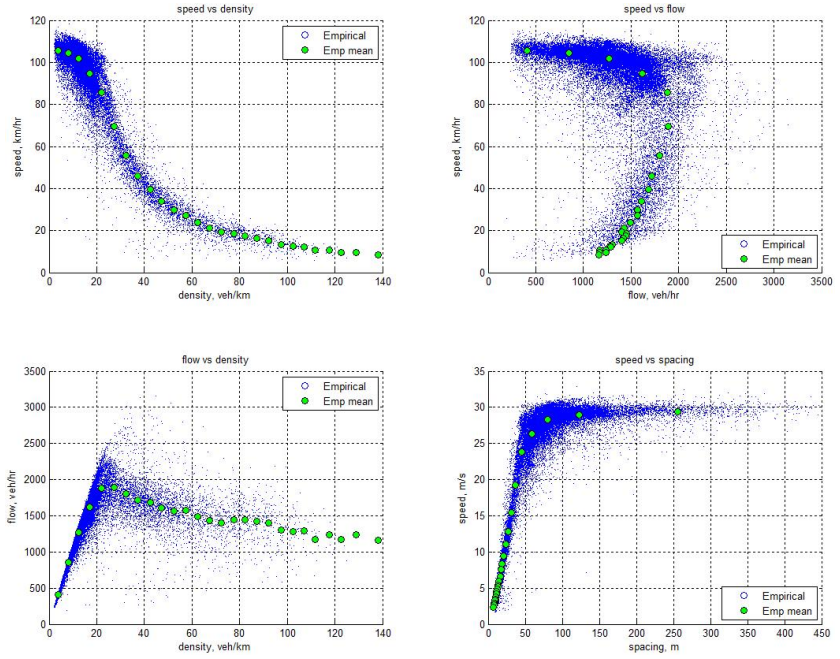


Figure 5.3: Observed q-k-v relationships (data source: NaviGAtor)

Noticeably, each subplot in the figure exhibits a trend which suggests a pair-wise relationship among flow, speed, and density, though such a relationship is of statistical significance. For example, the top-left subplot reveals a decreasing function 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 a freeway ($k \to 0$). Hence, one has the liberty to drive at his or her desired speed without being blocked by a slow driver ($v \to v_f$ the free-flow speed). The other intercept corresponds to a "Friday PM peak" scenario where everyone rushes home. As such, the road is jammed ($k \to k_j$ the jam density), resulting in a stop-and-go condition ($v \to 0$).

5.1.1 Greenshields model

Since the exact relationship between speed and density is unclear (and this is still true today!), Greenshields [50] proposed the use of a linear function to summarize the speed-density relationship. Such a function can be completely determined by knowing two points on the line: $(k = 0, v = v_f)$ and $(k = k_j, v = 0)$. Hence, the speed-density relationship (Figure 5.4) can be expressed as:

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

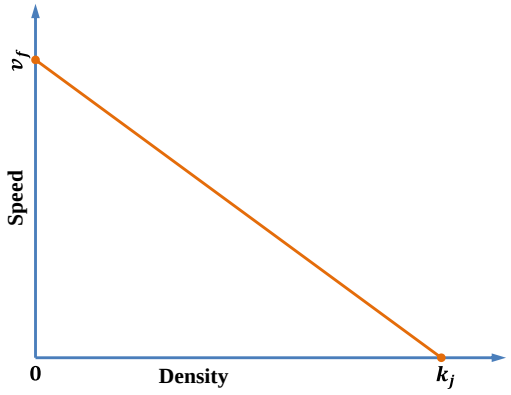


Figure 5.4: Greenshields speed-density relationship

This is the Greenshields model [50] which depicts a linear speed-density relationship. Combining the identity $q = k \times v$, one can derive the flow-density relationship implied by the Greenshields model:

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

In relation to the real world, the above model suggests the following: when density is close to zero ($k \to 0$), flow drops to zero ($q \to 0$) since the road is almost empty; when the road is jammed ($k = k_j$) flow also becomes zero ($q = 0$) because no one can move. In addition, since this is a quadratic function with a negative second order term, the corresponding $q - k$ curve must be parabolic with a downward opening, see the bottom left subplot of Figure 5.5. Therefore, starting from the origin ($k = 0, q = 0$), flow increases as density increases. This trend continues till, at some point ($k = k_m$), flow peaks ($q = q_m = \frac{v_f k_j}{4}$). After this point, flow begins to drop as density continues to increase and flow becomes zero ($q = 0$) when density reaches jam density ($k = k_j$). In this notation, q_m is the maximum flow, i.e. capacity, and k_m the optimal density, i.e. the density when flow peaks.

Similarly, one can eliminate k from the Greenshields model by using the identity and obtain a flow-density relationship:

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

This is again a quadratic function with an opening to the left, see the top right subplot of Figure 5.5. When flow is close zero ($q \to 0$), two scenarios are possible: (1) the road is so empty that the very few vehicles on the road are able to move at free-flow speed ($v \to v_f$) or (2) the road is jammed so that no one is able to move ($v \to 0$). Actually, entering the equation with a given flow value will result in two speed values: a lower one which corresponds to a congested condition and a higher one corresponding to a stable condition. When flow reaches capacity ($q = q_m$), the two speeds become one which is called the optimal speed, v_m.

Figure 5.5 summarizes the above discussion graphically and puts speed-density, flow-density, and speed-flow relationships together. Note that flow, speed, and density and their relationships are common notions in many fields such as fluid mechanics, electric engineering, and social science. However, relationships among these traffic flow characteristics such as those depicted in

Figure 5.3 and modeled in Figure 5.5 are unique to traffic flow and not observed in other kinds of flows. Hence, the model and its associated graphical representation that summarizes the pair-wise relationships among traffic flow characteristics are the distinguishing features of traffic flow. Therefore, they are referred to as the *fundamental diagram* in traffic flow theory. The work by Greenshields depicted in Figure 5.5 constitutes the first fundamental diagram in traffic flow theory.

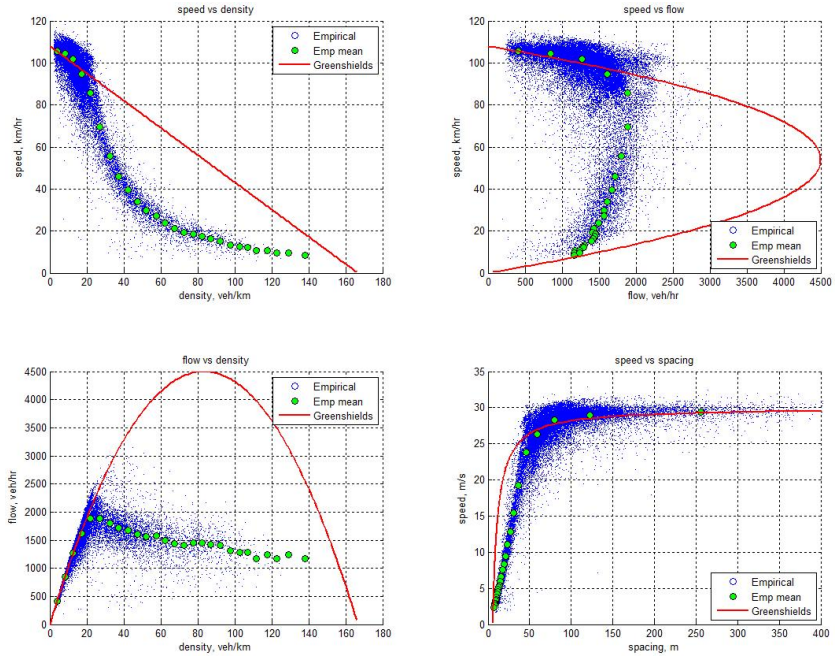


Figure 5.5: Fundamental implied by Greenshields model

Note that the three pair-wise relationships, i.e. 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 drivers' speed choices to the concentration of vehicles in the neighborhood of these drivers. In addition, the relationship features a one-to-one correspondence between speed and density, i.e. given a density, the relationship predicts a speed unambiguously. Therefore, the relationship is typically used in traffic flow theory to understand how drivers adjust their speeds in response to traffic in their vicinity. As can be seen later, the flow-density relationship is convenient

in 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 familiar with traffic engineering immediately realizes that the speed-flow relationship is extensively used to assist traffic engineering studies such as highway capacity analysis and determining level of service (LOS) of freeways and multilane highways.

5.1.2 Other single-regime models

Due to its simplicity and elegance, the Greenshields model, together with its associated fundamental diagram, is ideal for illustration purpose. Empirical observations reveal that the model lacks accuracy. For example, the model predicts that capacity ($q = q_m$) occurs at half jam density ($k_m = \frac{1}{2}k_j$). If an average vehicle length of 21 feet is assumed, the jam density will be somewhere around $5280/21 \approx 250$ vehicles per mile (vpm) and a half of this number is 125 vpm. However, field observations suggest that k_m is most likely in the range of 40-60 vpm. In addition, unlike the way that speed decreases linearly with density, field observations suggest that free-flow speed can be sustained up to a density of about 20 vpm before a noticeable speed drop can be observed.

Much efforts have been devoted subsequently to the formulation of speed-density models that approximate empirical observations closer than Greenshields. Table 5.1 is an incomplete list of these models.

Table 5.1: Single regime models

Author	Model	Parameters
Greenshields [50]	$v = v_f(1 - \frac{k}{k_j})$	v_f, k_j
Greenberg [49]	$v = v_m \ln(\frac{k_j}{k})$	v_m, k_j
Underwood [146]	$v = v_f e^{-\frac{k}{k_m}}$	v_f, k_m
Northwestern [29]	$v = v_f e^{-\frac{1}{2}(\frac{k}{k_m})^2}$	v_f, k_m
Drew [31]	$v = v_f[1 - (\frac{k}{k_j})^{n+\frac{1}{2}}]$	v_f, k_j, n
Pipes-Munjal [120, 97]	$v = v_f[1 - (\frac{k}{k_j})^n]$	v_f, k_j, n

where v_f is free-flow speed, k_j jam density, v_m optimal speed, k_m optimal

density, and n an exponent.

The above models share one thing in common - they are one-equation models, meaning the models apply to the entire range of density. Hence, these models are called single-regime models. Figure 19.4 shows the performance of these single-regime models by plotting them on top of empirical observations.

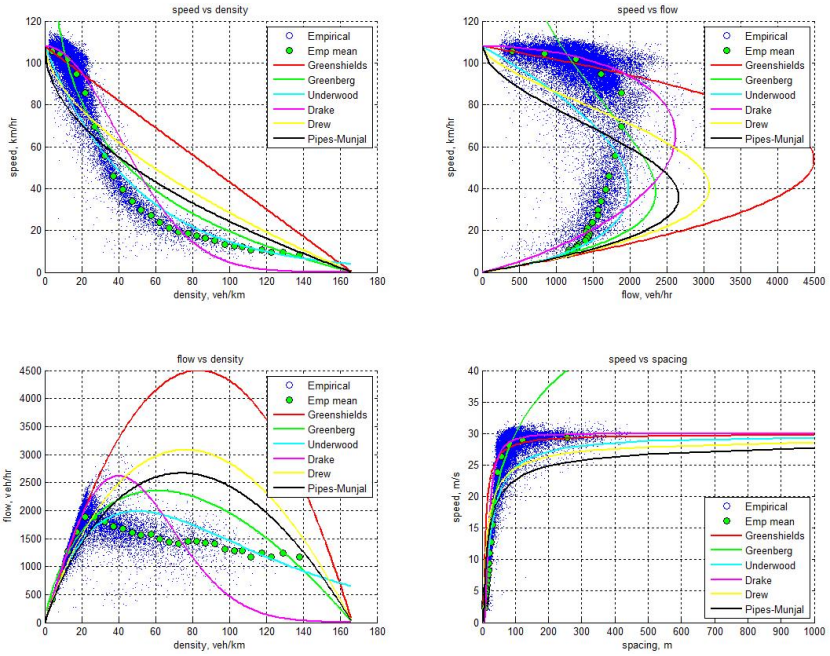


Figure 5.6: Comparison of single-regime models

5.2 Multi-regime models

It seems none of these single regime models is able to fit empirical observations reasonable well over the entire density range. Some models are good in one density range, while others are good at another. 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 piece-wise manner using multiple equations. This gives rise to multi-regime models, an incomplete list of which (including Edie [34], two-regime linear, modified Greenberg, and three-regime models) is presented in Table 5.2.

Table 5.2: Multi-regime models

Regimes Models	Free-flow	Transitional	Congested
Edie model	$v = 54.9e^{-k/163.9}$ $k \leq 50$	- -	$v = 26.8\ln(162.5/k)$ $k > 50$
2-regime model	$v = 60.9 - 0.515k$ $k \leq 65$	- -	$40 - 0.265k$ $k > 65$
Modi. Greenberg	$v = 48$ $k \leq 35$	- -	$v = 32\ln(145.5/k)$ $k > 35$
3-regime model	$v = 50 - 0.098k$ $k \leq 40$	$v = 81.4 - 0.913k$ $40 < k \leq 65$	$v = 40 - 0.265k$ $k > 65$

May [88] presented a comparison of these multi-regime models. Similar work illustrated in Figure 5.7 is found in [152].

5.3 Can we go any further?

Though all relationships presented above take deterministic forms, the actual relationships are essentially statistical. For example, a speed-density relationship predicts that, when density is $k = 20$ vpm, speed will be $v = 60$ mph. However, in reality, observed speed may vary over a wide range (e.g. 42, 53, 67, etc.) and these observations form a distribution. The significance of these models lies in their ability to predict a value that makes statistical sense. For example, if one observes traffic sufficiently long and collects enough speed samples, the likelihood of having a speed value in the neighborhood of 60 mph is very high. Figure 19.4 illustrates 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 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 5.8. Empirical observations seem to support such a proposition. For example, Figure 5.9 plots the observed mean and standard deviation of

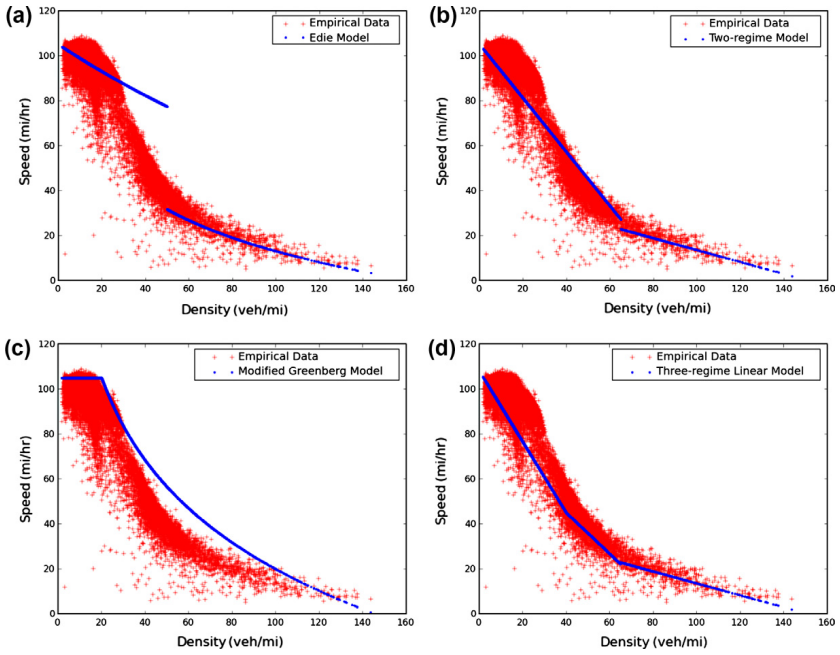


Figure 5.7: Comparison of multi-regime models

speed-density relationship in a single figure. Hence, the deterministic speed-density relationship in the form of

$$v = f(k)$$

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 references [153, 79, 152] for details of an attempt of such a stochastic speed-density relationship.

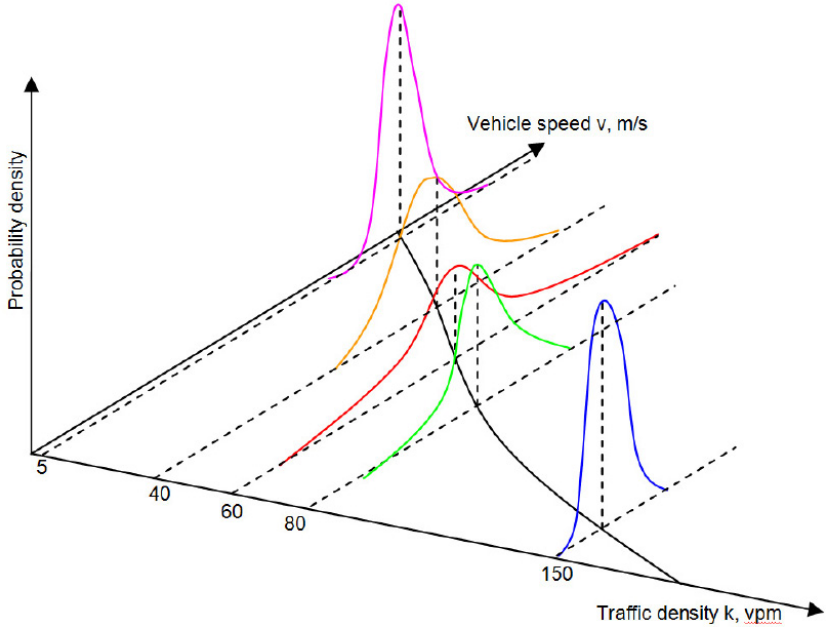


Figure 5.8: 3D representation of speed-density relationship

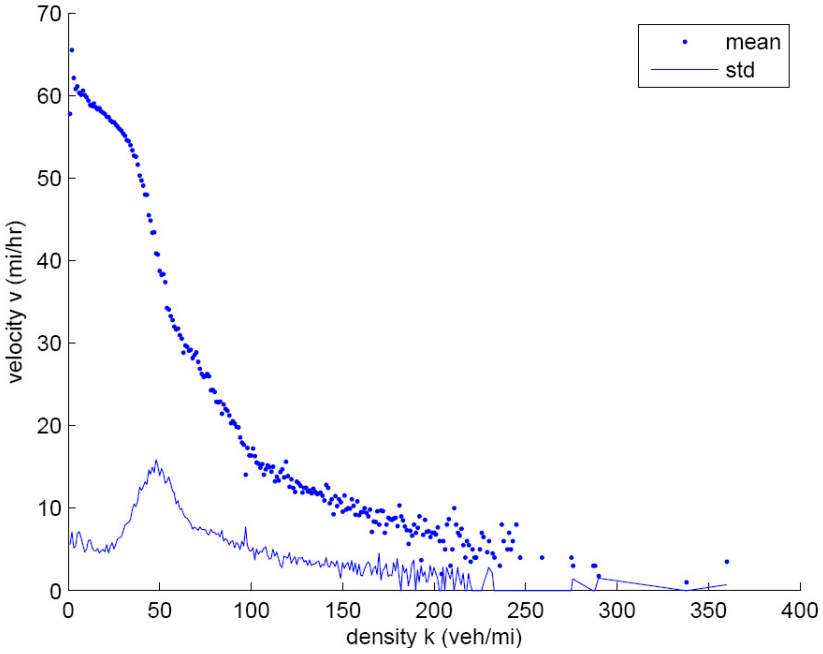


Figure 5.9: Mean and variance of speed-density relationship

Part II

Macroscopic Modeling

Chapter 6

Conservation Law

In previous chapters, two types of relationships among traffic flow characteristics have been discussed:

The identity

$$q = k \times v$$

The following is noted about this relationship: (1) it is an identity, i.e., it is self-guaranteed by generalized definition of traffic flow characteristics; (2) it is location- and time-specific, i.e. $q(x,t) = k(x,t) \times v(x,t)$, i.e., flow, speed, and density must refer to the same location and time.

The pair-wise relationships

$$v = V(k)$$

$$q = Q(k)$$

$$v = U(q)$$

A few comments on these relationships follow: (1) they define a fundamental diagram and hence differentiate vehicular traffic flow from other kinds of flows; (2) they are location-specific, i.e. different locations and roads may have different underlying fundamental diagrams; (3) they are equilibrium models, i.e. they describe a steady-state behavior in the long run, and hence not specific to a particular time; (4) they are deterministic, but such functions are only of statistical significance, i.e. the equal signs do not strictly

hold in the real world. Based on (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 applying 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 bottlenecks, if any?
- In case of congestion, how long does it last and how far do queues spill back?
- If an incident occurs, what is the best strategy to clean up so that the impact on traffic is minimized?

Answers to these questions involve analyzing dynamic change of traffic states over time and space. Unfortunately, the above relationships or models are only capable of describing traffic states without providing 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 examining a small volume of roadway traffic as a continuum. Here traffic flow is treated as a one-dimensional compressible fluid such as gas. Laws of conservation apply to this kind of fluid and the first-order form of conservation is mass conservation, also known as the continuity equation.

6.1 The continuity equation

There are several ways to derive the continuity equation, each takes a different perspective on the small volume of roadway traffic.

Derivation I: finite difference

The following derivation is found in reference [42]. Suppose a highway section is delineated by two observation stations at x_1 and x_2 and the segment length is $\Delta x = x_2 - x_1$. 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:

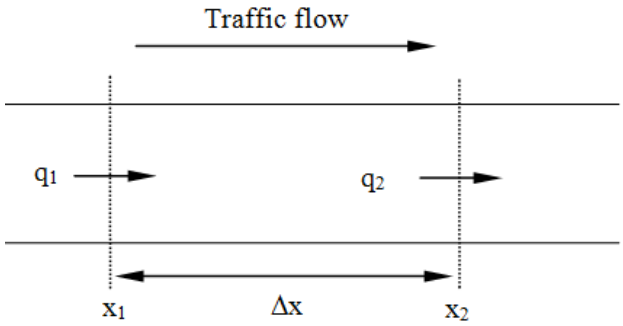


Figure 6.1: Deriving continuity equation I

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

The change of vehicles in the section is

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

Assume 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 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 can not be created or destroyed inside the section, the change of vehicles should be the same in the same section during the same time interval. Therefore, $\Delta N = \Delta M$, i.e.

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

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

Divide both sides by $\Delta x \Delta t$,

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

Let $\Delta x \to 0$ and $\Delta t \to 0$, 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 6.2 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. Based on vehicle conservation, the following relationship holds:

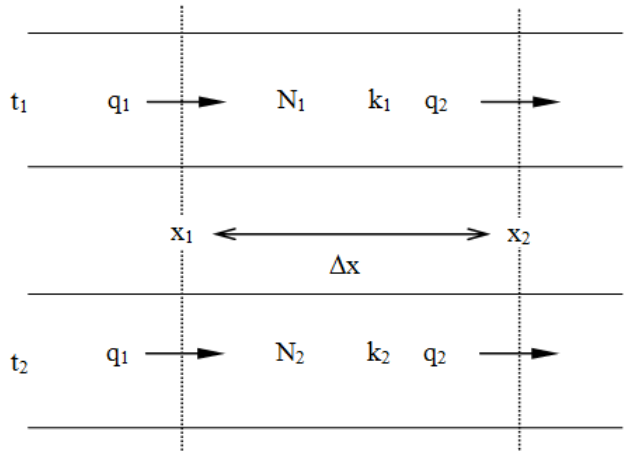


Figure 6.2: Deriving continuity equation II

Vehicles at t_2 = vehicles at t_1 + inflow during Δt - outflow during Δt

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 $\Delta x\Delta t$

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

Let $\Delta x \to 0$ and $\Delta t \to 0$,

$$q_x + k_t = 0$$

Derivation III: fluid Dynamics

Figure 6.3 illustrates a small fluid cube of size $\delta x \times \delta y \times \delta z$. Fluid velocity v and density k at two sides of the cube also are shown.

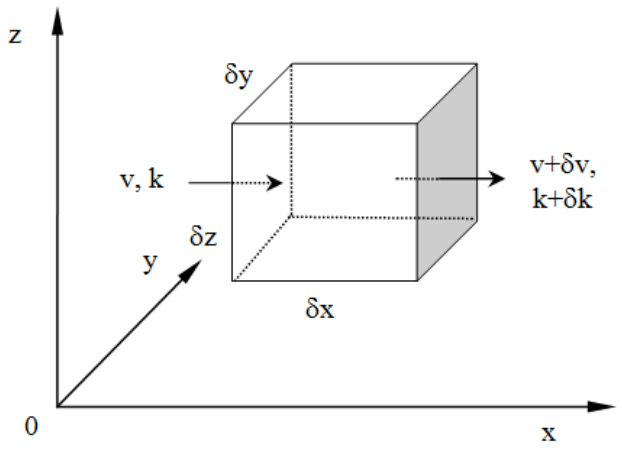


Figure 6.3: Deriving continuity equation III

Mass flows into the cube = $vk\delta y\delta z$

Mass flows out of the cube

$= (v + \delta v)(k + \delta k)\delta y \delta z$

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

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

Mass stored in the cube

= mass that flows in - mass that flows out

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

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

Ignore the high-order term,

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

$= \frac{\partial (kv)}{\partial x}\delta x \delta y \delta z$

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

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

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

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

Therefore

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

Highway traffic constitutes a special case of the above with only one dimension, see Figure 6.4. Using the result derived above, one obtains:

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

Note that $q = kv$. Therefore

$$q_x + k_t = 0$$

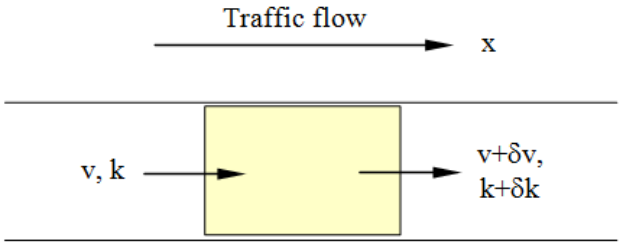


Figure 6.4: Reducing 3D to 1D

Derivation IV: scalar conservation law

This derivation is adopted from reference [66]. Consider a cell in the time-space domain bounded by $(x_1, x_2) \times (t_1, t_2)$. Let traffic flow, speed, and density be functions of time and space, i.e. $q = q(x,t)$, $v = v(x,t)$, and $k = k(x,t)$. Obviously, the conservation of vehicles in the cell requires the following:

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

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

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

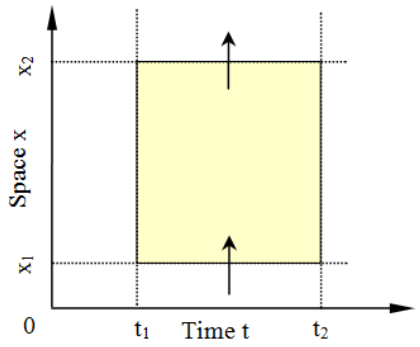


Figure 6.5: Deriving continuity equation IV

$$\int_{x_1}^{x_2}\int_{t_1}^{t_2}\frac{\partial k(x,t)}{\partial t}dtdx = -\int_{t_1}^{t_2}\int_{x_1}^{x_2}\frac{\partial q(x,t)}{\partial x}dxdt$$

$$\int_{x_1}^{x_2}\int_{t_1}^{t_2}[\frac{\partial k(x,t)}{\partial t}+\frac{\partial q(x,t)}{\partial x}]dxdt = 0$$

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

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

i.e.

$$q_x + k_t = 0$$

Derivation V: 3D representation of traffic flow

As discussed in Chapter 4 3D representation of traffic flow, the surface which represents the cumulative number of vehicles, N, can be expressed as a function of time t and space x, i.e. $N = N(x,t)$. Density at time space point (x,t) is the first partial derivative of $N(x,t)$ with respect to x, but takes negative value:

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

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

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

If both flow and density have first-order derivatives, i.e.

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

and

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

Therefore

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

i.e.

$$q_x + k_t = 0$$

6.2 First-order dynamic model

Traffic evolution is about 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 (e.g. $k_0 = k(x,0)$)and boundary conditions (e.g. $q(t) = q(x_0,t)$). One recognizes know that traffic states are functions of time t and space x and the continuity equation have been derived to describe the dynamic relation between flow $q(x,t)$ and density $k(x,t)$:

$$q_x + k_t = 0$$

This equation involves two independent variables x and t and two dependent variables $q(x,t)$ and $k(x,t)$. The equation is not solvable because the number of unknown variables are greater than the number of equations. One needs to supply another equation. In this case, the identity comes up:

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

Unfortunately, by adding one equation, a third unknown variable is introduced, i.e. speed $v(x,t)$. One needs to supply a third equation. Considering

that no better choice is available, an equilibrium traffic flow model (e.g. the Greenshields model) has to be employed. Such a model takes the form of:

$$v = V(k)$$

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

$$\begin{cases} q_x + k_t = 0 \\ q = kv \\ v = V(k) \end{cases} \tag{6.1}$$

Starting from initial and boundary conditions, the above system of equations is solvable. Consequently, one is able to find traffic state at an arbitrary time space point (x, t), i.e. $q(x, t)$, $v(x, t)$,and $k(x, t)$. Therefore, one is able to answer questions posed at the beginning of this Chapter.

Chapter 7

Waves

In order to solve the set of equations presented at the very end of Chapter 6, one has to leave the topic of traffic flow for a moment and study waves first. One would agree that this is necessary when looking at Figure 7.1 where vehicle trajectories recorded in the field are plotted in 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 have to be attributed to the knowledge of waves. This chapter serves to provide a jump start about waves.

Figure 7.1: Traffic waves observed on a highway

7.1 Wave phenomena

Waves are everywhere in the real world. When throwing a pebble into a pound, one sees ripples circling out. This is a wave, see Figure 7.2. When the audience at a football stadium becomes thrilled with rows of audience standing up and sitting 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 water, audience, and rope are media. Applying 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 7.1 are examples of such a wave.

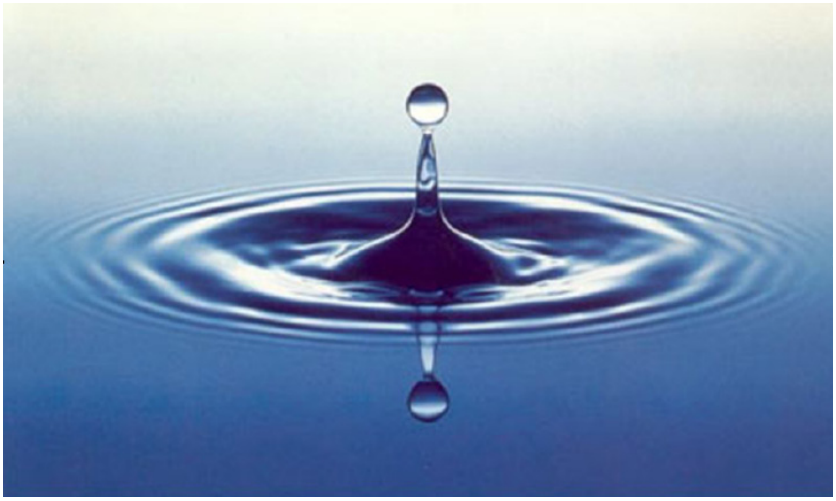


Figure 7.2: Surface waves

7.2 Mathematical representation

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

7.2.1 Notation

If a dependent variable, k, is a function of independent variables t and x, we denote $k = k(t, x)$ and 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 $k(x,t)$ is an equation that involves one or more partial derivatives of k with respect to t and x, For example:

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

7.2.2 Terminology

PDEs can be classified based on 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

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

Homogeneity

The above general first-order PDE is

- *Homogeneous* if $R(t,x,k) = 0$
- *Non-homogeneous* if $R(t,x,k) \neq 0$

Linearity

In the above general first-order PDE, if both P and Q are independent on k, i.e. $P = P(t, x)$, $Q = Q(t, x)$ and,

- if R is also independent on k, i.e. $R = R(t, x)$, then the PDE is *strictly linear*, e.g. $2xk_t + 3k_x = 5t$.
- if R is dependent on k, i.e. $R = R(t, x, k)$, then the PDE is *linear*, e.g. $2xk_t + 3k_x = 5k + 3$.
- if R is dependent on k in a nonlinear manner, then the PDE is *semi-linear*, e.g. $2xk_t + 3k_x = e^k$

In particular, if $P = P(t, x, k)$, $Q = Q(t, x, k)$, and $R = R(t, x, k)$, then the PDE is *quasi-linear*, e.g. $k_t + (3k + 2)k_x = 0$

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

Now, test yourself by classifying the following PDEs:

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

7.3 Traveling waves

Many PDEs have solutions in a traveling wave form $k(t, x) = f(x - ct)$ [1]. Figure 7.3 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

[1] the following discussion is derived from [75]

profile at time 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.

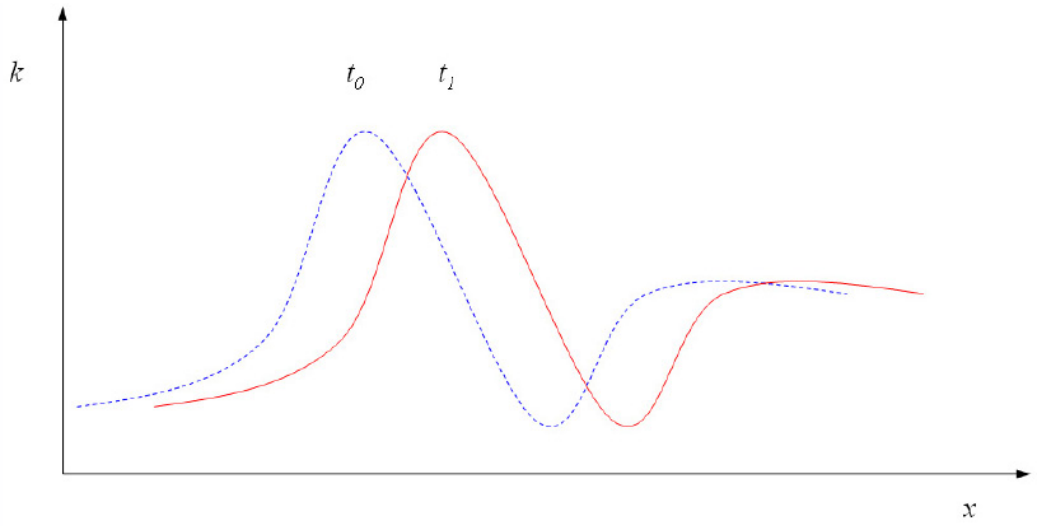


Figure 7.3: A traveling wave

7.4 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 take 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 $k_{xx} = f''$

Plug the above into the wave equation, one obtains

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

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

7.5 Wave Front and Pulse

A traveling wave is called a wave front if

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

A traveling wave is called a pulse if $k_1 = k_2$. See Figure 7.4.

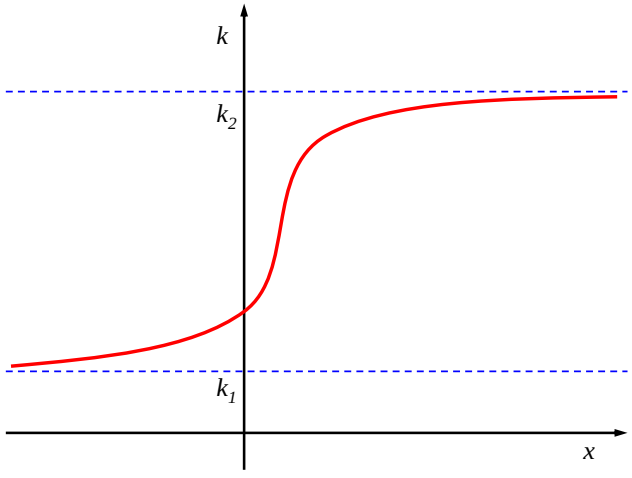


Figure 7.4: Wave front and pulse

7.6 General solution to wave equation

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 right hand side terms is a traveling wave, their superposition may not necessarily be.

Example 1

Solve the following wave equation with initial conditions (IC)

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

Applying the above general solution and after some math, one obtains

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

Example 2

Solve the following wave equation with initial conditions (IC)

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

Applying the result in Example 1 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

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

7.7 Characteristics

Consider Example 1, since $f(x) = k(x,0)$ and $g(x) = k_t(x,0)$, the solution can be transformed to the following form:

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

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 initial condition at points $(0, x^* - ct^*)$ and $(0, x^* + ct^*)$ and the interval **I** bounded by the two points (inclusive), i.e. $I = [x^* - ct^*, x^* + ct^*]$. This is illustrated in the left part of Figure 7.5. Therefore, the interval **I** if called the *domain of dependence* of point (t^*, x^*).

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 7.5.

Characteristics

Notice that, in the left part of Figure 7.5, the two lines emitting from point (t^*, x^*) intersecting 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* (not to be mixed up with traffic flow characteristics).

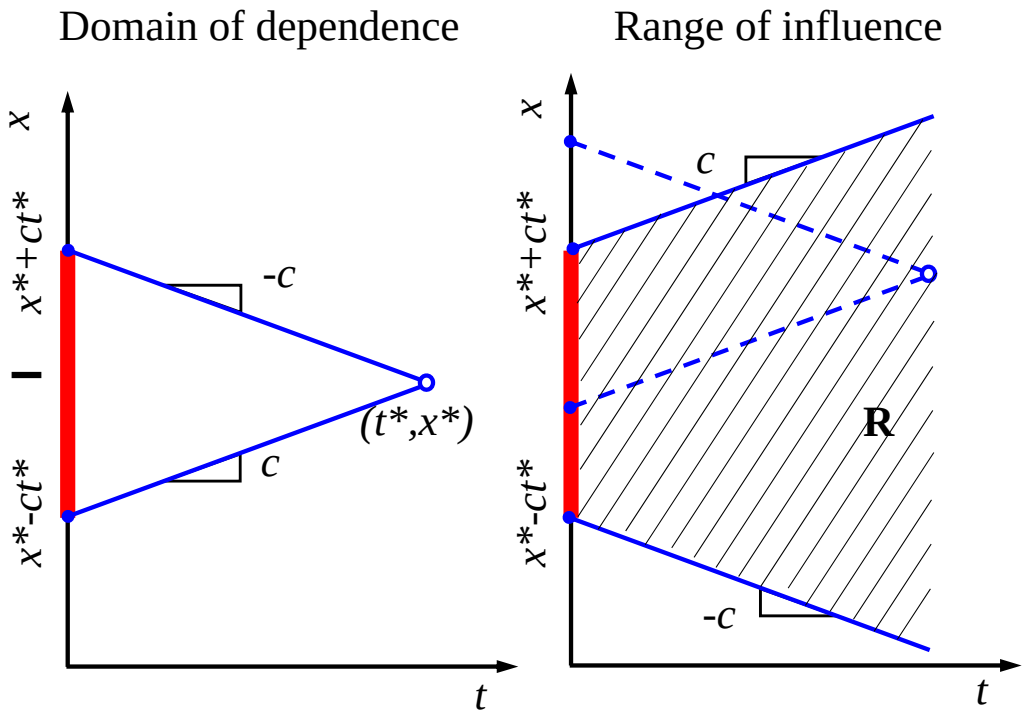


Figure 7.5: Characteristics

7.8 Solution to Wave Equation

In a special case where $k_t(0, x) = 0$, the solution of the wave equation in Example 1 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, the solution k at (t,x) is constructed by taking average of $k(0,x_1)$ and $k(0,x_2)$, see the right part of Figure 7.5.

Example 3

Use characteristics to solve the following wave equation:

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

In this equation, the traveling wave speed $c = \pm 2$, i.e. $k(t,x) = f(x \pm 2)$. First, one constructs an xt-plane. locate points 0 and 1 on x-axis. Then one emits two characteristics (their slopes are ± 2) from each of the two points, see Figure 7.6. The four characteristics partition the xt-plane into six regions as labeled. Take an arbitrary point (t_0, x_0) for example. The solution at this point is found my emitting two characteristics from this point. Then find the intersections of the two characteristics on 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, i.e. $k(t_0, x_0) = \frac{1}{2}$.

Using similar technique, the solution in other regions can be determined. To sum up, the solution to the above wave equation is the following:

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

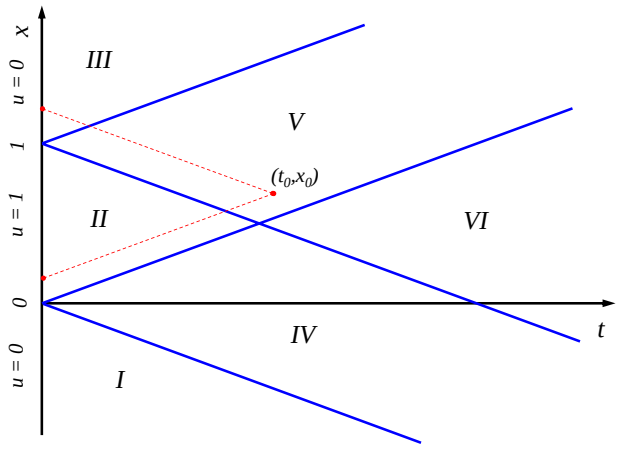


Figure 7.6: Examples of Characteristics

Chapter 8

Waves and Conservation Law

Discussion in Chapter 7 presents the following notion:

1. For some wave equations such as Example 3 in Chapter 7, 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 emitting two lines, called characteristics, from (t, x) at slopes c and -c.

3. These characteristics intersect 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 $k(t, x) = \frac{1}{2}[k(0, x_1) + k(0, x_2)]$.

8.1 Method of characteristics

Now let us consider a very simple PDE derived from the conservation law with initial condition. In Chapter 6, the conservation law leads 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 a 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 is be clear shortly) and the new goal is to find the k value at an arbitrary point $(t, x(t))$ on the curve, i.e. $k(t, x(t))$. To find the solution, let us examine how k changes along the curve $x = x(t)$. The rate of change of k in time is the first (and total) derivative of k with respect of time t, i.e.:

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

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, i.e. 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 a slope of c. Therefore, the equation of the line is found by solving the following ordinary differential equation:

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

which yields:

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

At time $t = 0$, this line intersects 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 to all points on this line. Such a line is called a *characteristic*. Sounds familiar? Yes, it bears the same meaning as the characteristic in Chapter 7 where it is a line emitting from a time-space point with slope c which is the speed of the traveling wave $f(x - ct)$.

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

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

Example 1

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

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

Soultion: following the above procedure, one obtains:

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

8.2 Some properties

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

8.2.1 Properties of characteristic

In the above example, the characteristic is a straight line and this is so because c is a constant. Similarly, another characteristic emitted from another time-space point is also a straight line. In addition, the two straight lines are parallel since the have the same slope c. Figure 8.1 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 $k(x,t)$ is not necessarily flat. A *kinematic wave* is a family of characteristics which carry and propagate signals, such as those characteristics illustrated in the figure.

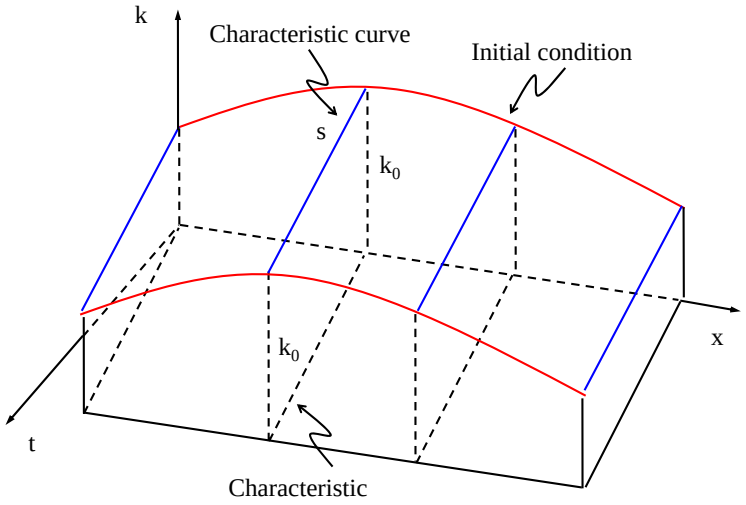


Figure 8.1: Illustration of parallel characteristics

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

Example 2

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

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

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 intercept x_0. Consequently, two characteristics may intersect. See Figure 8.2 for an illustration.

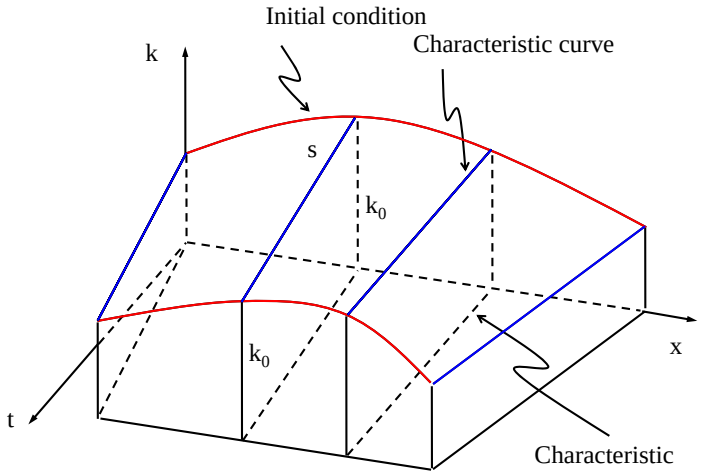


Figure 8.2: Illustration of non-parallel characteristics

Example 3

In this example, c explicitly depends on x and or t, e.g. $c = t$. 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 a parabola. In addition, characteristics emitted from different time-space points are no longer parallel. Instead, they may intersect.

Therefore, the following comments are noted about characteristics:

- 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 multi-valued at the intersection of two characteristics. Such an occurrence is called a *gradient catastrophe*.

8.2.2 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, i.e.

$$k_t + ck_x = 0$$

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 decreasing at the rate of 1, i.e. $k = k_0 - t$ where k_0 is found in initial conditions. Figure 8.3 illustrate such a case.

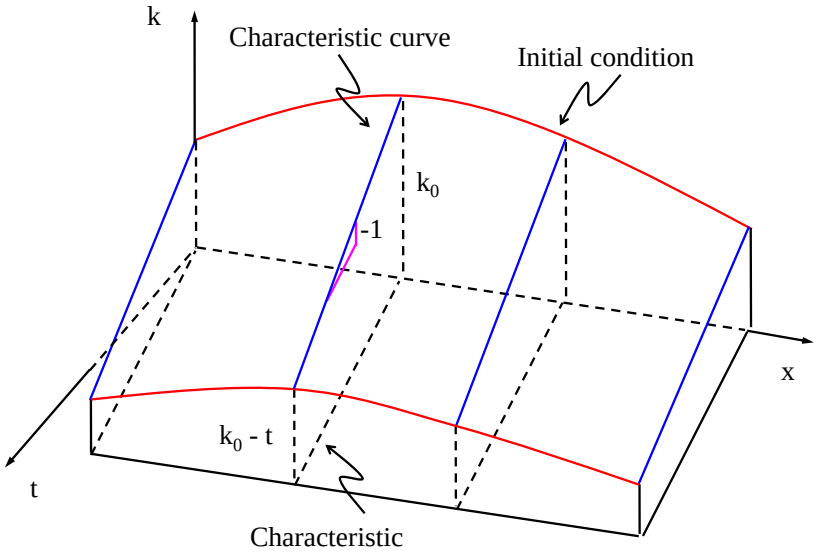


Figure 8.3: Solution of a non-homogeneous PDE

Chapter 9

Shock and Rarefaction Waves

In the previous chapter, the method of characteristics was discussed as a means to solve continuity equation (i.e. conservation law) with initial condition [1]:

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

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

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

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

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

[1] Discussion in this chapter is mainly derived from reference [75]. Readers are encouraged to read the original book for in-depth information.

9.1 Gradient catastrophes

In the above example, if c is a constant, characteristics emitted 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 emitted from two time-space points are still straight lines but not necessarily parallel, in which case they may intersect and the solution at this intersection may be multi-valued. For example, Figure 9.1 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. When the two characteristics intersect at point C, the solution profile will have infinity gradient at this point. The formation of such an infinity gradient is called a *gradient catastrophe* and the time when infinity gradient occurs is called the *break time* t_b. Figure 9.2 presents a few frames of time development of the solution profile. Notice that the dot on the top 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 can be a family of characteristics moving farther and farther apart without any intersection, see Figure 9.3, such a wave is called an *expansion wave*. The corresponding time development of solution profile is shown in Figure 9.4. 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.

9.2 Shock waves

Continuing the above discussion, if two characteristics intersect, the solution at the intersection will be multi-valued. However, if one allows discontinuity at the intersection, it is possible to construct a piece-wise smooth solution. For example, Figure 9.5 illustrates such a solution where curve $x_s(t)$ in

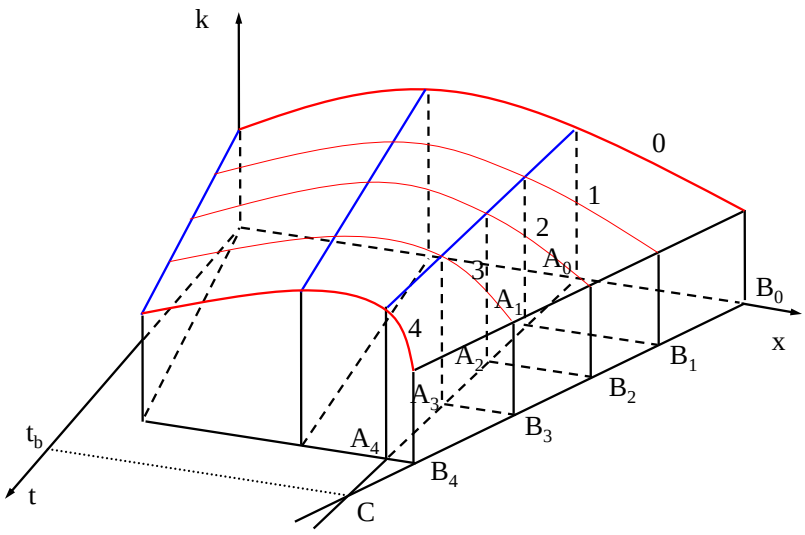


Figure 9.1: A gradient catastrophe

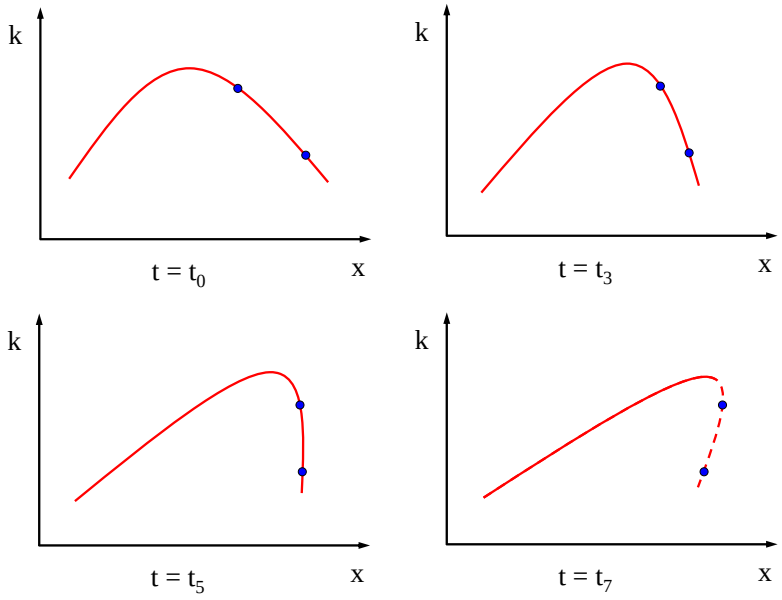


Figure 9.2: Top of profile overtakes bottom of profile

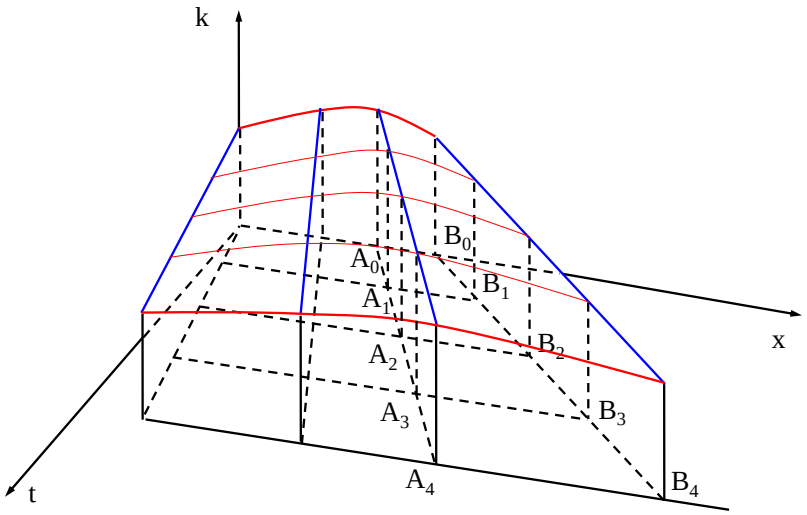


Figure 9.3: Characteristics farther and farther apart

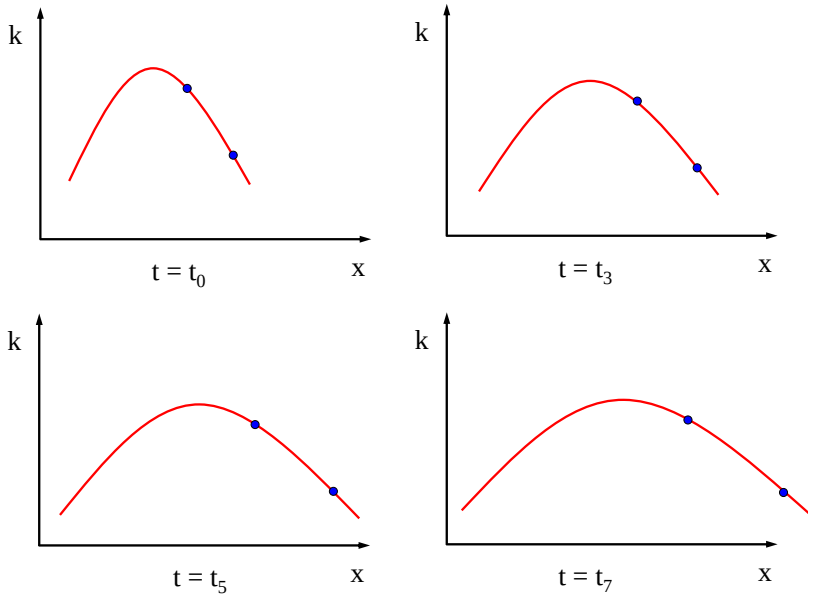


Figure 9.4: Solution profile thinned out

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 piece-wise smooth solution of the PDE is called a shock wave solution.

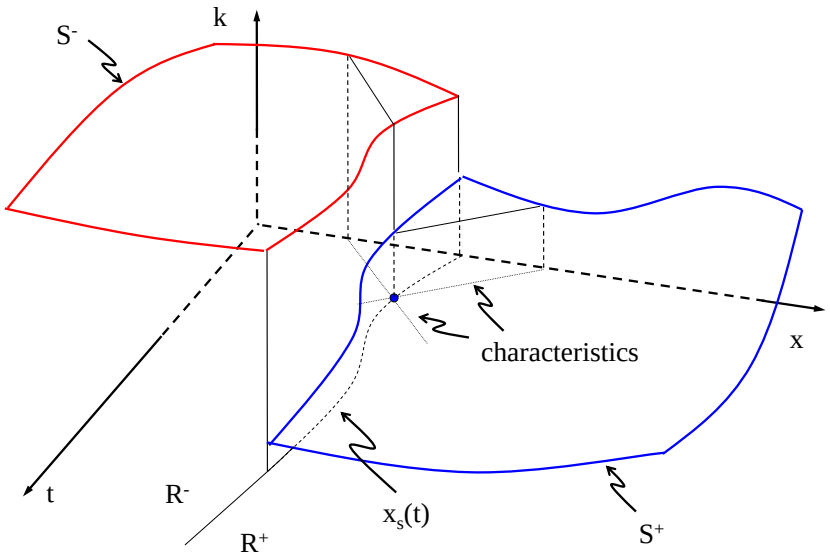


Figure 9.5: Piece-wise solution - shock wave

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 is the location along which a shock wave forms, such a curve is called a *shock path*. In Figure 9.6, 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-Hugonoit jump condition*:

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

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

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

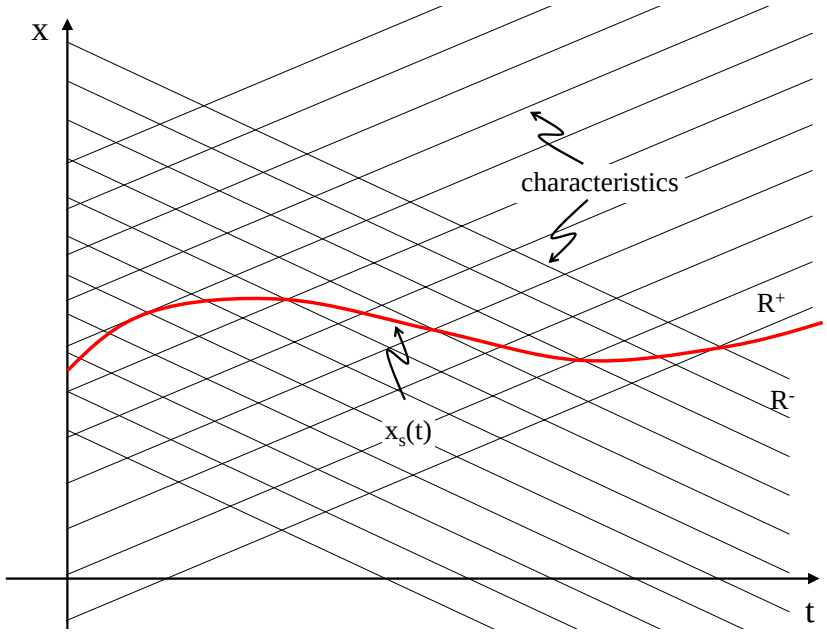


Figure 9.6: Shock path

Example

Solve the conservation law with following initial conditions:

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

Obviously, characteristics emitted below $x = 0$ are straight, parallel lines with slope 1. These characteristics carry the same constant value of $k = 1$ and, hence, $q = 1/2k^2 = 1/2$. Similarly, characteristics emitted above $x = 0$ are horizontal lines (i.e. slope is 0). They carry $k = 0$ and, hence, $q = 0$. The origin is a known point on the shock path. According to the Rankine-Hugonoit jump condition, the slope of the shock path is

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

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

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

Therefore, the solution is

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

The solution is illustrated in Figure 9.7. 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, a shock wave separates two kinematic waves with an abrupt change of k value; a shock path is the projection of shock locations onto x-t plane.

9.3 Rarefaction waves

If the initial condition in the above example reverses, i.e.

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

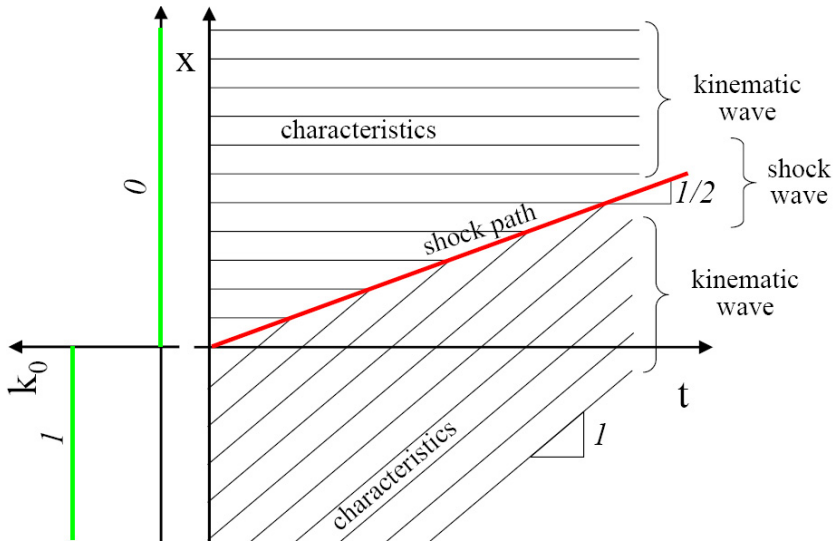


Figure 9.7: An example of shock path

characteristics of this PDE should be drawn like Figure 9.8. 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 relax the step function of the initial condition by assuming that k_0 smoothly varies from 0 to 1 over a small distance Δx, see Figure 9.9, the slopes of characteristics emitted between Δx will gradually increase from 0 to 1 so that any point in the solution space is swept by one and only one characteristic.

In order to return to the step function of the initial condition, one takes the limit $\Delta x \to 0$, so Figure 9.9 reduces to Figure 9.10. Now the empty area is filled with a fan of characteristics emitted from the origin. If one cuts the solution space with a few planes $t = t_0, t_1, t_2, ...$ 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 9.11. 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.

The rarefaction wave can be used to construct a solution for the following

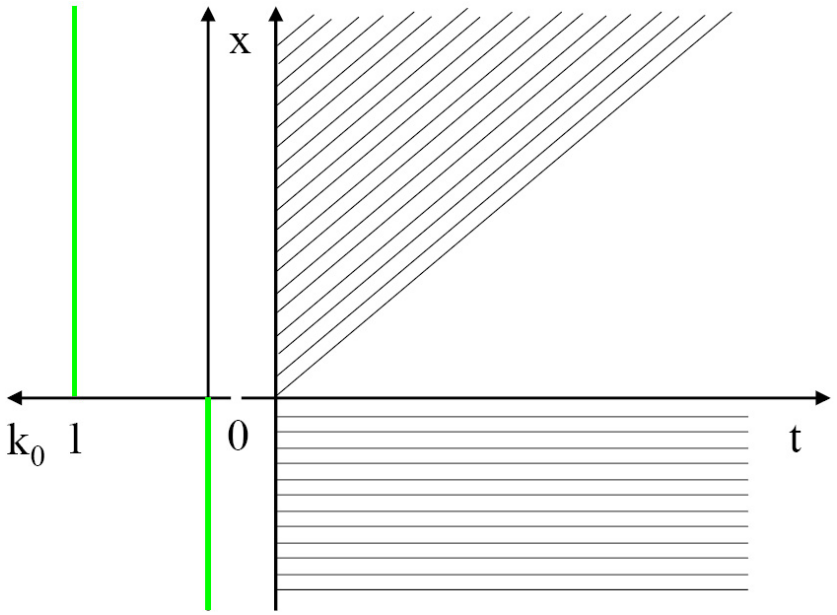


Figure 9.8: Characteristics without an intersection

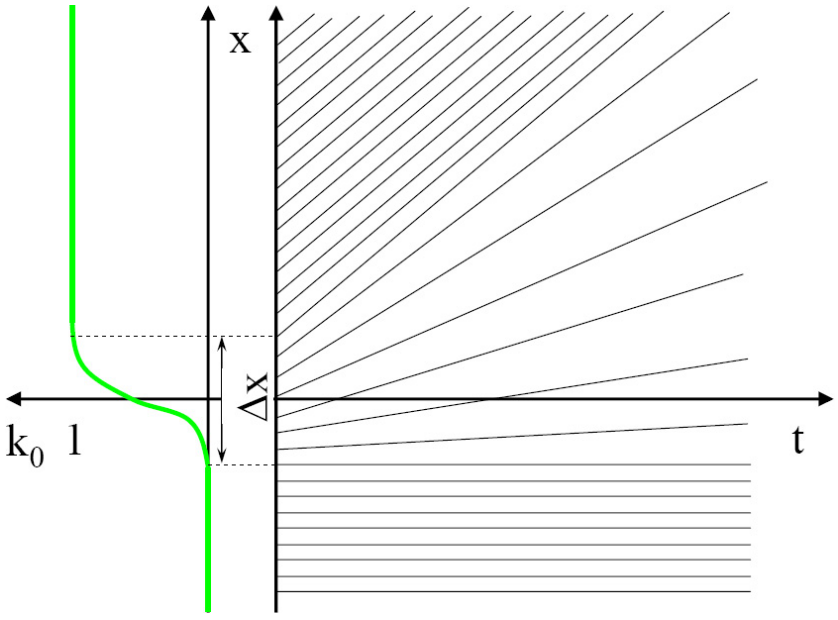


Figure 9.9: Filling an empty area with characteristics

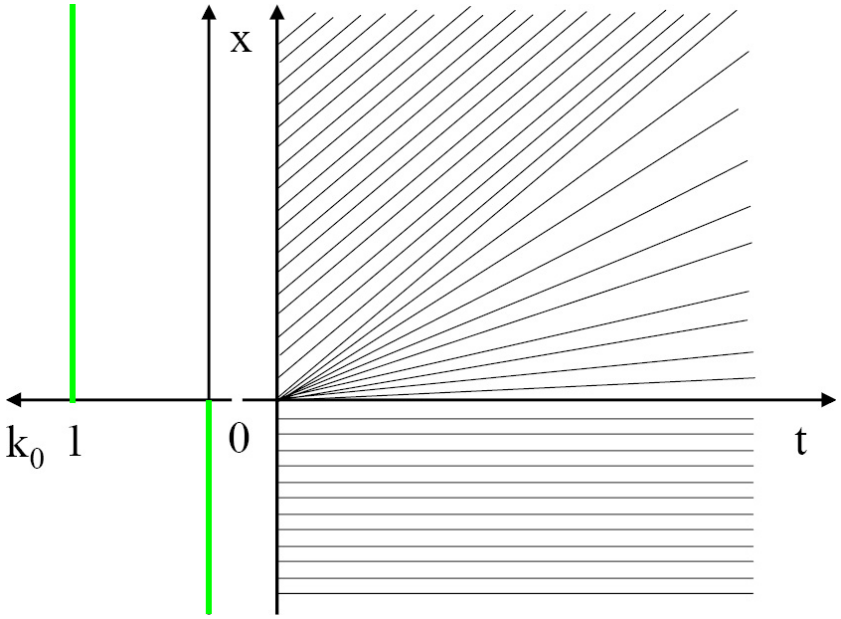


Figure 9.10: A rarefaction wave

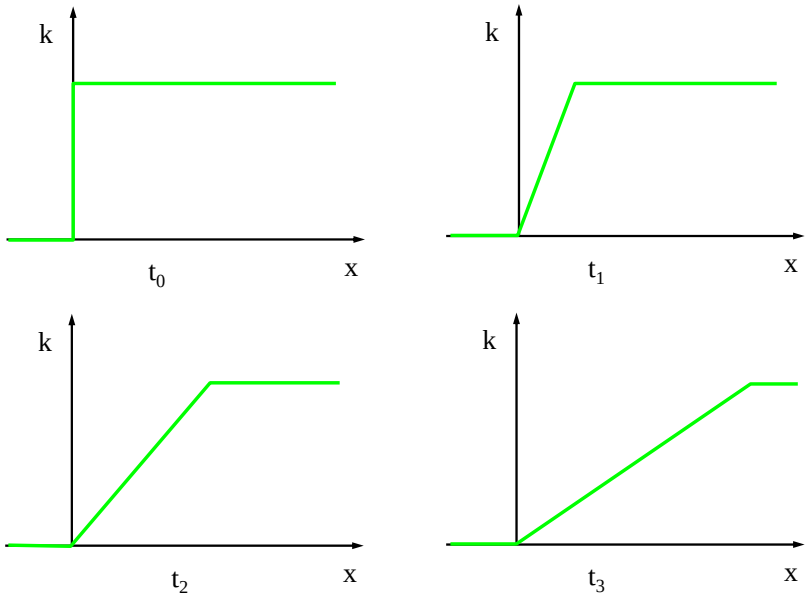


Figure 9.11: Time development of the rarefaction wave

conservation law problem:

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

From the initial condition and using the method of characteristics, solutions to the two areas swept by the two parallel characteristics in Figure 9.8 can be easily determined:

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

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,

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

The conservation law can be re-written as

$$k_t + kk_x = 0$$

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

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

Solving this equation gives $f'(x/t) = 0$ or $f = \frac{x}{t}$. If $f'(x/t) = 0$, $f(x/t) = k(x,t) = a$ where a is an integral constant. A simple check along $x = 0$ and $x = t$ reveals that this solution does not satisfy the Rankine-Hugonoit jump condition. Hence, this solution is not a shock wave solution. The smooth solution is given by $k(x,t) = f(x/t) = x/t$. Hence, the rarefaction wave solution of the original problem is:

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

Meanwhile, it is also possible to construct a shock wave solution by choosing a constant a such that $0 < a < 1$ and applying the Rankine-Hugonoit jump condition:

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

Figure 9.12 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 multiple. 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?"

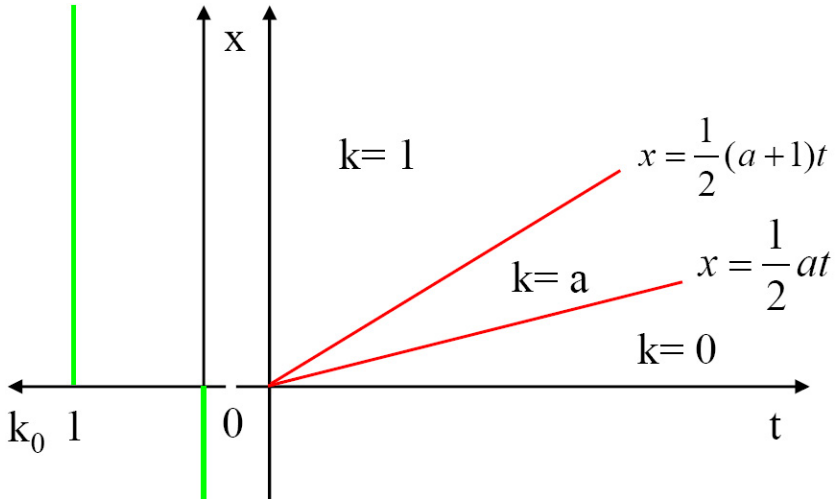


Figure 9.12: A shock wave solution

9.4 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:

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

for $\Delta x > 0$ and $t > 0$ and such a condition is shown in Figure 9.13.

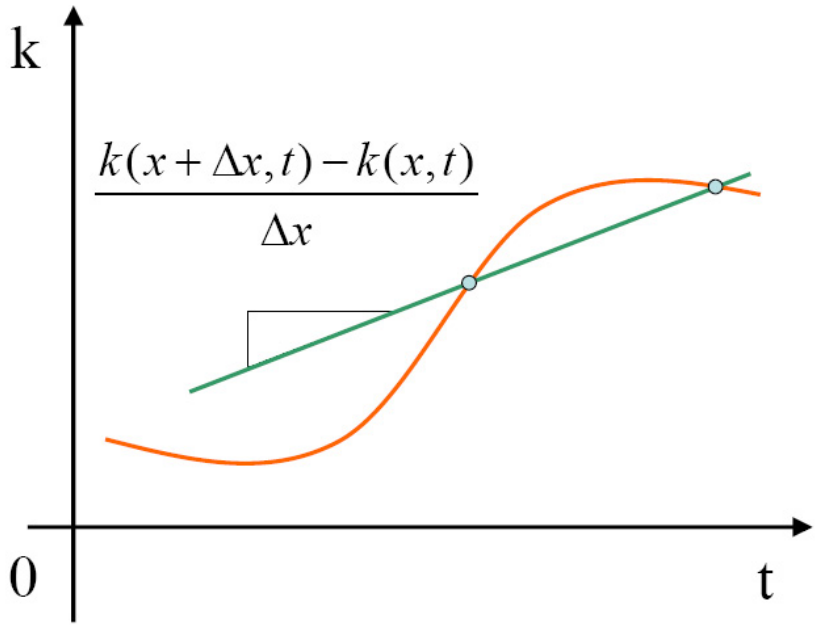


Figure 9.13: Entropy condition

Now, let us check if a shock wave solution satisfies this condition. This is done by slicing the solution space in Figure 9.12 using a plane AA' and project the result onto the $k-x$ plane, as shown in Figure 9.14. The solution profile consists of three discontinuous sections with $k = 0, a, 1$. If one choose two arbitrary points with 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 could not find a positive constant E to satisfy the entropy condition and, hence, shock wave solutions do not make physical sense.

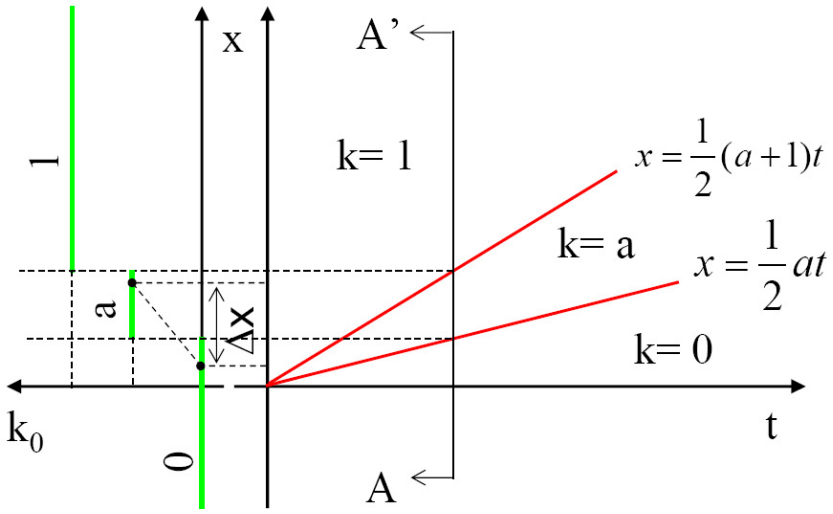


Figure 9.14: Entropy condition in a shock wave solution

To apply the same technique to check the rarefaction solution, one slices the solution space with a plane at AA' and the resultant solution profile is illustrated in Figure 9.15. 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 most physical sense:

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

As an excise, find rarefaction wave solutions to the following PDEs:

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

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

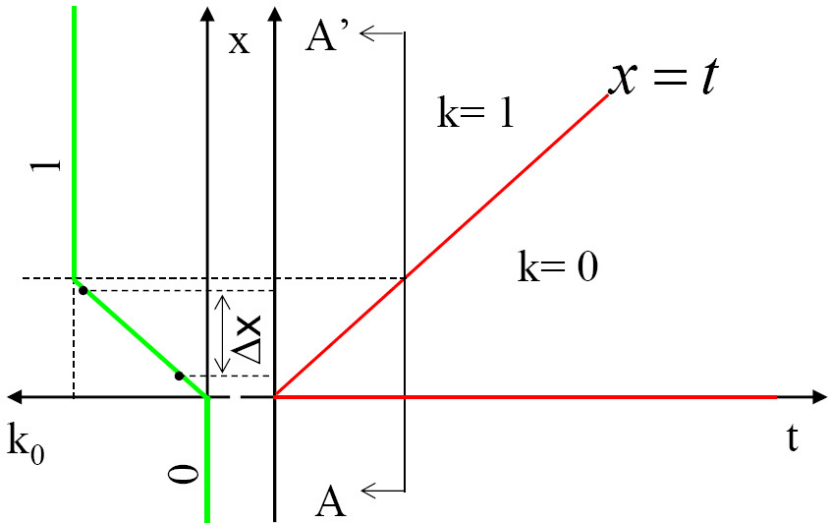


Figure 9.15: Entropy condition in the rarefaction solution

9.5 Summary of wave terminology

- 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 chock 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 diverging characteristics which cause two neighboring signals to move farther and farther apart which, in turn, causes a rarefaction wave.

Chapter 10

LWR Model

In previous chapters, we temporarily left traffic flow and concentrated on the conservation law (Chapter 6), waves (Chapter 7), solutions to the conservation law (Chapter 8), and shock waves (Chapter 9). The purpose of these chapters is to pave the road to addressing traffic dynamics and unveiling traffic evolution on highways.

10.1 The LWR model

At the end of Chapter 6, A dynamic traffic flow model was formulated based on conservation law:

$$\begin{cases} k_t + q_x = 0 \\ q = kv \\ v = V(k) \end{cases} \tag{10.1}$$

where $q = q(x,t)$ is flow, $k = k(x,t)$ is density, $v = v(x,t)$ 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:

$$\begin{cases} k_t + q_x = 0 \\ q = Q(k) \end{cases} \tag{10.2}$$

or further

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

where $Q'(k) = \frac{dQ(k)}{dk}$. This is the so-called LWR model [80, 130] just to honor the three pioneers, Lighthill, Whitham, and Richards, who originally studied this problem. Note that LWR model is essentially a first-order, homogeneous, quasi-linear partial differential equation.

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

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

2. Start with an arbitrary point on the x axis $(x^*, 0)$, determine the k value at this point $k_0(x^*)$ and the value of $c(x^*) = Q'(k_0(x^*))$.

3. Emit a straight line s from point $(x^*, 0)$ 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 $k(x_s, t) = k_0(x^*)$.

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. In case that a characteristic has multiple intersections, use Rankine-Hugonoit jump condition to determine the right intersection. Repeat this step and find out adjacent intersections. Connect these intersections to form a shock path. The solution at both sides of the shock path should be piece-wise smooth with a jump along the shock path which forms a shock wave.

6. In case that 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 solutions, apply the entropy condition to select a solution that makes the most physical sense.

8. After following the above steps, 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 (x, t) is of interest, one simply follows its characteristic all the way back to the x axis and read $k_0(x)$ off the initial condition. This $k_0(x)$ is the k value at the time-space point in question. Consequently, one finds the corresponding $q(x, t) = Q(k(x, t))$ and $v(x, t) = \frac{q(x,t)}{k(x,t)}$. Hence the solution $k(x, t)$, $q(x, t)$, and $v(x, t)$ of any time-space point (x, t) can be determined.

Note that conservation law (and consequently 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 based on 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 using them as a look-up table is the last resort when no better choice is available.

10.2 Example: LWR with Greenshields model

Greenshields model [50] assumes the following linear v-k relationship:

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

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

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

Hence

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

If the parameters are given as follows: $v_f = 60$ miles per hour (mph) and $k_j = 240$ vehicles per mile (vpm). The explicit form of the LWR model becomes

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

Find solutions at points ($t = \frac{1}{2}$ hour, $x = 25$ mile) and ($t = 1$ hour, $x = 65$ mile) using the following initial condition:

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

Following the above solution procedure, one constructs a time-space diagram, shows the initial condition aside the diagram, and identifies the two points in question, see Figure 10.1. Next, one constructs characteristics. All characteristics emitted between $0 < x \leq 10$ will bear a k value of 40 which can be read from the initial condition, so the slope of these characteristics is $c = 60 - \frac{k}{2} = 40$. Point $(t = \frac{1}{2}, x = 25)$ is within this area and the characteristic passing this point intercepts x-axis at $(0, 5)$. Hence, $k(\frac{1}{2}, 25) = k(0, 5) = 40$. Similarly, All characteristics emitted from $x > 10$ have slope $c = 50$ and point $(t = 1, x = 65)$ is within this area. The characteristic passing this point intercepts x-axis at $(0, 15)$. Hence, $k(1, 65) = k(0, 15) = 20$.

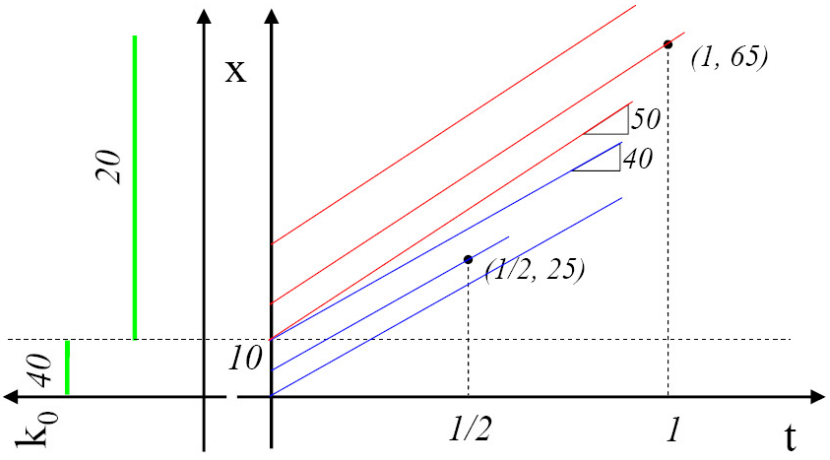


Figure 10.1: Example: LWR with Greenshields model

10.3 Shock wave solution to 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 which is the speed of the fast kinematic wave. Similarly, the speed of the slow kinematic wave is 40. 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 reverses, i.e. the slow platoon leads the fast platoon, soon 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 after, the second vehicle in the fast platoon will have to slow down, and the third vehicle, the fourth vehicle, and so on. The "slowing down" effect will propagate backward along the fast platoon. The propagation of 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 xt-plane is called a *shock path*.

As discussed in the method of characteristics, a characteristic carries a constant k value (i.e., density), the intersection of two characteristics will inevitably have two k values. This means that at this point two traffic conditions co-exist and, after the intersection, the two platoons resume their original conditions along their respective characteristics. This situation does not make any physical sense. In order to develop a solution that is physically meaningful, one has to make the solution piece-wise 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 jump (or shock) at the intersection.

To illustrate the idea, the previous example is revisited with the fast platoon being behind. On x-t plane in Figure 10.2, two families of characteristics, i.e. two kinematic waves, are drawn, but this time those characteristics emitted between $0 < x < 10$ will have a slope of 50, while those emitted from $x > 10$ have slope 40. Since the fast kinematic wave is behind, it will catch up with the slow kinematic wave, i.e., 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 pla-

toon, i.e. all points in this region carry the condition of the slow platoon, and the other region belongs to the fast platoon, i.e., all points in this region carry the condition of the fast platoon. When one moves across the shock path, traffic condition changes suddenly from one to another, i.e. experiencing a shock - that is how the term shock wave derives its name. Therefore, it is convenient to read from the figure that $k(\frac{1}{2}, 25) = k(0, 0) = 20$ and $k(1, 65) = k(0, 25) = 40$.

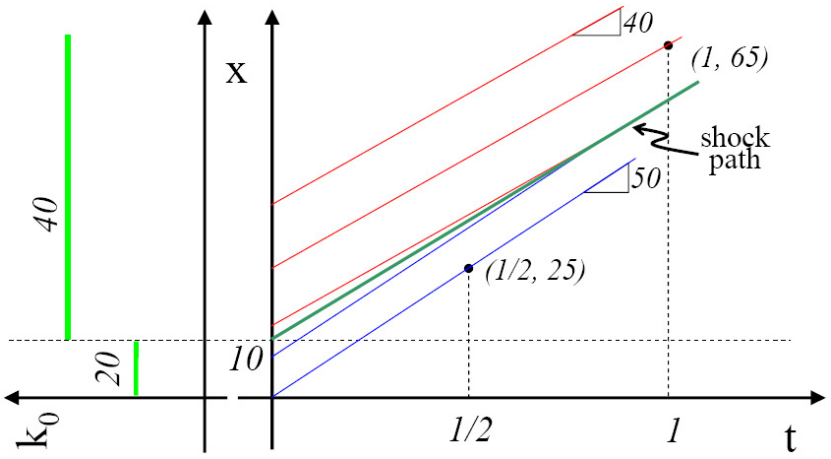


Figure 10.2: Example: Revisit LWR with Greenshields model

10.4 Riemann problem

In the above example, two properties are noticeable:

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

As discussed in Chapter 8, if c is a constant or dependent on k but not explicitly on t or x, the resultant characteristic is a straight line which is the case in the above example since $c = Q'(k) = v_f - 2\frac{v_f}{k_j}k$.

Still in Chapter 8, 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 piece-wise constant k_0, each family of characteristics will have the same slope, i.e. they are parallel. The above example is such a case.

As discussed in Chapter 9, the slope of a shock path is determined by the Rankine-Hugonoit jump condition. If the initial condition consists of piece-wise constant k_0, then solutions k and q at both sides of the shock path are piece-wise constnat. Consequently, the Rankine-Hugonoit jump condition will result in a shock path with a constant slope, i.e. 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 is given as piece-wise constant. In general, conservation law with piece-wise constant initial data is referred to as a *Riemann problem*, named after Bernhard Riemann who was a German mathematician.

10.5 LWR model with general q-k relationship

In the above examples, the underlying q-k relationship is explicitly given, e.g., 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 Greenshields model suffers inaccuracy and, often times, the underlying q-k relationship is graphically given by fitting from empirical data. In this case, solution to the LWR model with general q-k relationship is typically determined graphically.

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

$$\begin{cases} k_t + q_x = 0 \\ q = Q(k) \\ k(t,0) = k_0(x) = \begin{cases} A & \text{if } x \leq 0, \\ B & \text{if } x > 0. \end{cases} \end{cases} \tag{10.3}$$

where the underlying q-k relationship is given in Figure 10.3 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 aside. Since this

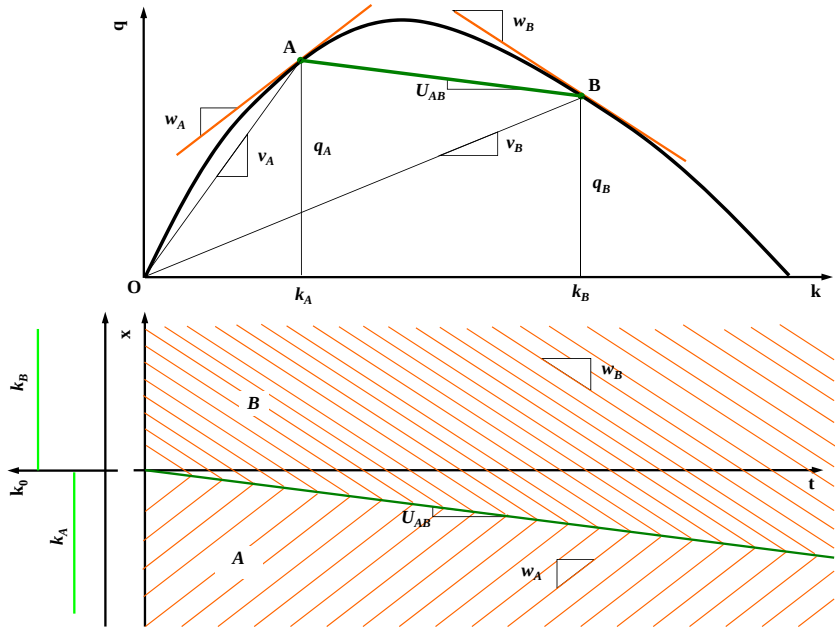


Figure 10.3: Example: LWR with general q-k relationship

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 from $x > 0$. Since the speed of kinematic wave A is

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

i.e. 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 emitted 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 B represents a heavy, slow platoon in front and A a light, fast platoon behind. A will catch up with 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-Hugonoit 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 time-space diagram. Therefore, the shock path can be determined 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 which consists of two piece-wise 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 condition A (q_A, k_A, u_A).

10.6 Shock path and queue tail

In Figure 10.3, the shock path actually represents the time-varying location which separates the fast platoon and the slow platoon, i.e. the tail of a moving queue. As the leading vehicle of the fast platoon catches up 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 10.4 shows a few snapshots to illustrate such a dynamic process.

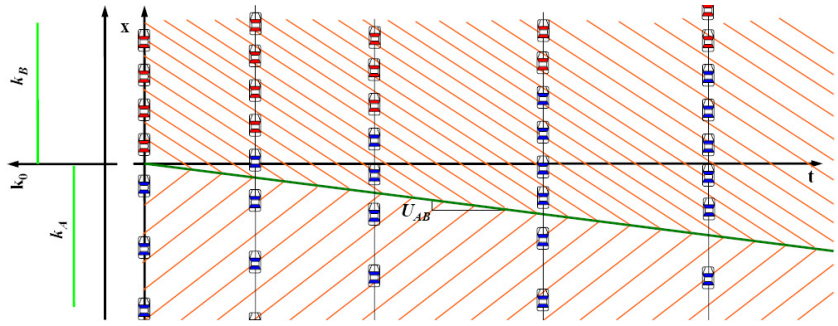


Figure 10.4: Shock path and queue tail

One may have recognized that, though characteristics are used to illustrate how to find the shock path, they are actually unnecessary. With a known point on the shock path and shock speed, the shock path can be

determined directly without drawing characteristics. In the above example, the solution can be directly constructed 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.

10.7 Properties of 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 10.5 gives the full picture.

10.7.1 Flow-density relationship and speeds

Traffic speed v If operating point A is known which represents a traffic condition with flow q_A and density k_A, 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 origin O and operating point A.

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

$$v_f = \lim_{A-O} v_A = \lim_{k_A \to 0} \frac{q_A}{k_A}$$

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 catches up with traffic with condition B.

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

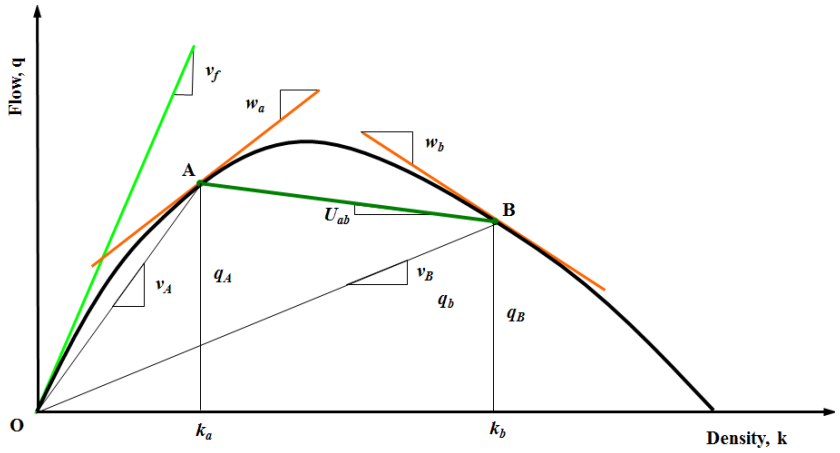


Figure 10.5: Speeds in a flow-density relationship

10.7.2 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, i.e. everything is relative to an observer standing stationary at 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 what 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 emitting a line from the origin with w_A as its slope. Then run a vertical line through point A intersecting the emitted 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. For another example, suppose traffic is operating at condition B which is at the congested side of the q-k curve. The kinematic wave speed now is 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 emitting a line from origin O with slope w_B which slants downward. Run a vertical line through point B intersecting the emitted 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.

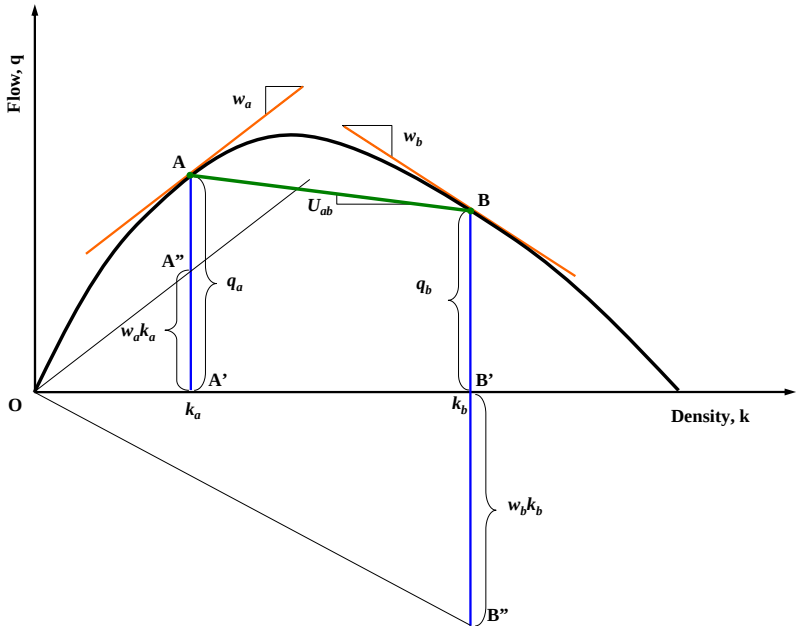


Figure 10.6: Traffic flow observed by a moving observer

Chapter 11

Example LWR Problems

The previous chapter has focused on LWR model and its solutions using the method of characteristics and shock waves. The purpose of this chapter is to supply readers with a few more involved examples to apply this methodology to solve some traffic flow problems.

11.1 A bottleneck with varying traffic demand

Traffic arriving at the upstream of a highway was initially at condition A, see Table 11.1 and Figure 11.1. At 9:00AM, the condition switches to condition B. After one hour, the condition switches back to condition A. The capacity at the bottleneck is 1400 veh/hr. Find how far the queue extends back and how long the queue persists.

Table 11.1: Traffic data: a bottleneck with varying traffic demand

Condition	q (veh/hr)	k (veh/mi)	v (mi/hr)
A	600	8.57	70
B	2000	40	50
D	1400	21.5	65
D'	1400	130	10.8

Solution

With the aid of graphical construction in Figure 11.1, the rate at which the queue grows is:

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

The queue tail extendes back at this rate for one hour, so the farthest point it reaches is 6.67 mi upstream of the bottleneck. The rate at which the queue dissipates is:

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

So the time needed to dissipate the queue is $\frac{6.67}{6.60} = 1.01$ hr and the total time that the queue persists is 2.01 hour.

11.2 An intersection with constant demand

Traffic arrives at a signalized intersection at a constant rate of 800 veh/hr and all conditions are given in Table 11.2. The intersection is under pretimed signal control with cycle length 90 seconds and a split of effective green/red of 0.5/0.5. Determine the farthest point of the queue.

Table 11.2: Traffic data: an intersection with constant demand

Condition	q (veh/hr)	k (veh/mi)	v (mi/hr)
A	800	25	32
C	1600	80	20
D	0	180	0
O	0	0	40

Solution

With the aid of graphical construction in Figure 11.2, when the signal turns red, a queue grows at a rate of

$$U_{AD} = \frac{q_D - q_A}{k_D - k_A} = \frac{800 - 0}{25 - 180} = -5.16 \text{ mi/hr}$$

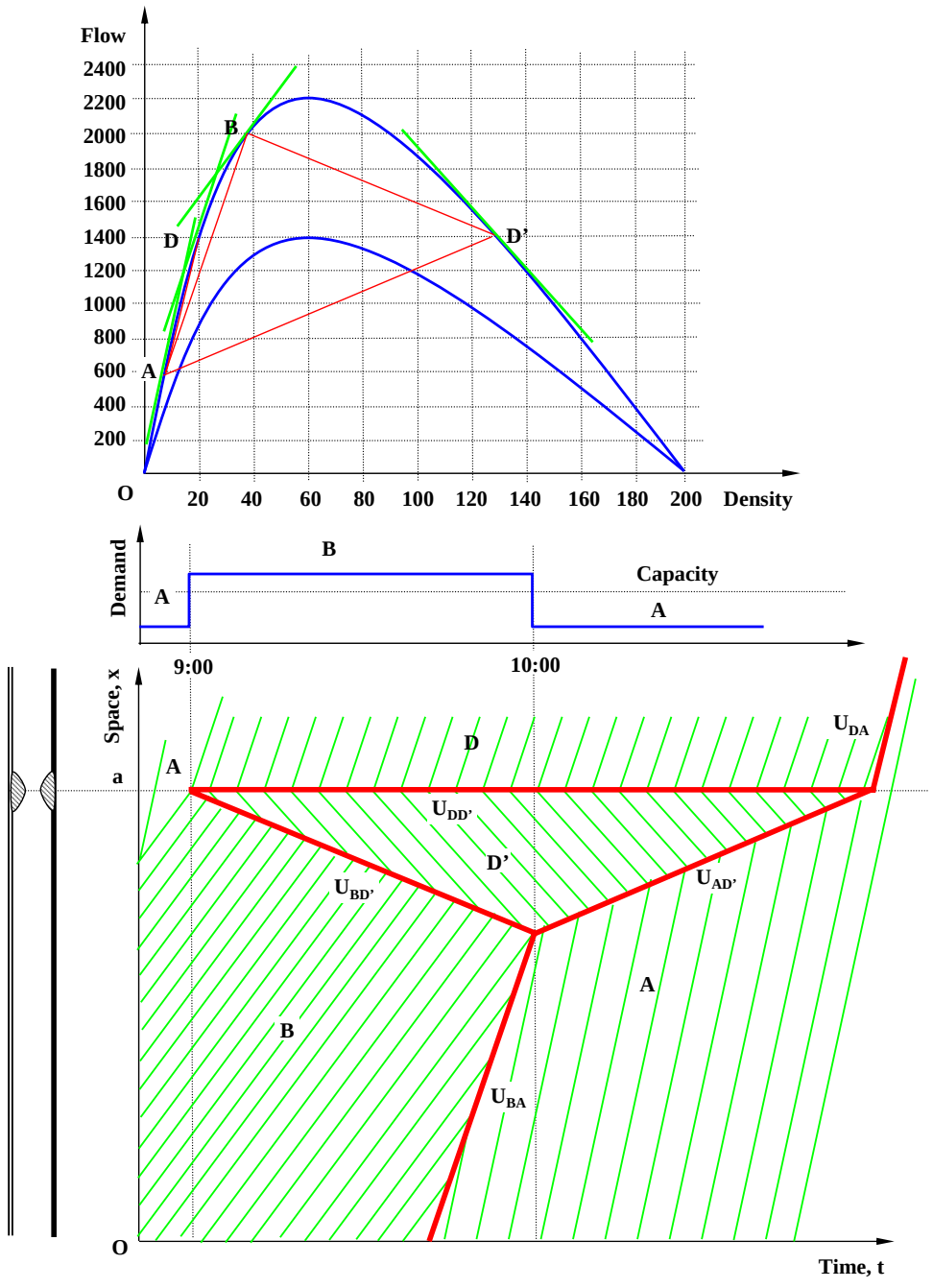


Figure 11.1: A highway bottleneck with varying traffic demand

When the signal turns green, the queue shrinks at a rate of

$$U_{DC} = \frac{q_C - q_D}{k_C - k_D} = \frac{0 - 1600}{180 - 80} = -16 \text{ mi/hr}$$

In triangle a-d-c: $U_{AD} = \frac{cd}{ca}$, $ca = \frac{cd}{U_{AD}}$
In triangle b-d-c: $U_{DC} = \frac{cd}{cb}$, $cb = \frac{cd}{U_{DC}}$
$ca - cb = 45$ s

$$\frac{cd}{5.16} - \frac{cd}{16} = \frac{45}{3600}$$

$cd = 0.095$ mi $= 503$ ft

So the queue extends back 503 ft from the intersection.

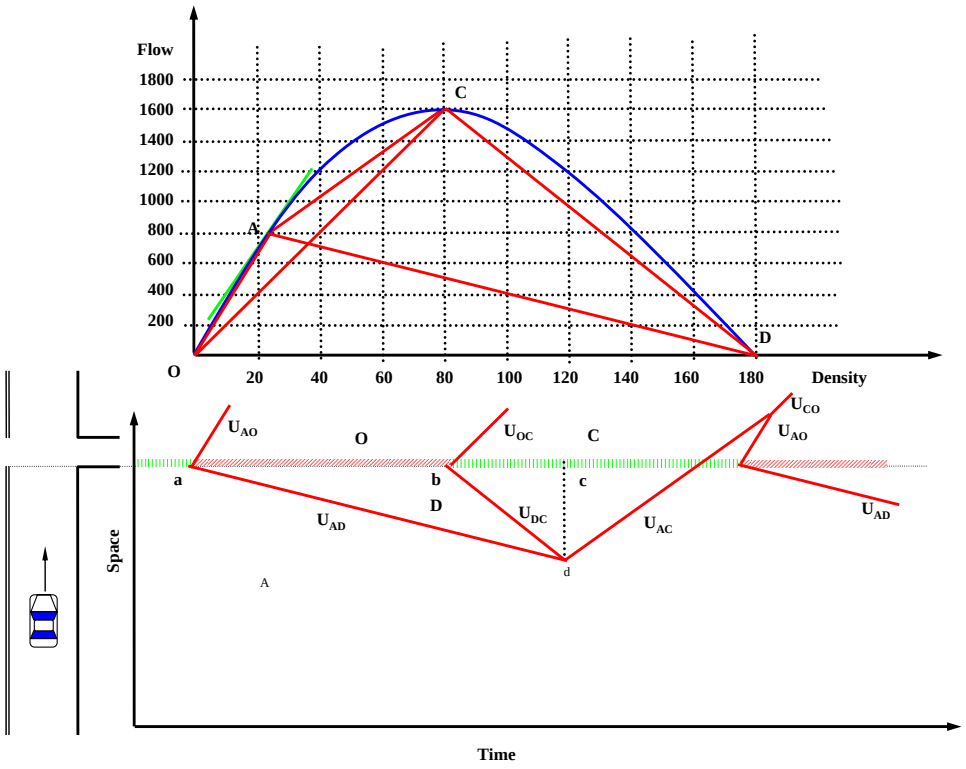


Figure 11.2: An intersection with constant demand

11.3 A moving bottleneck

A freeway was initially operating under condition A, see Table 11.3. At 2:30 pm, a sluggish truck entered the freeway traveling at a speed of 13.3 mi/hr. The truck turned off the freeway at the next exit 6.67 mi apart. Find when the impact of the truck will disappear.

Table 11.3: Traffic data: a moving bottleneck

Condition	q (veh/hr)	k (veh/mi)	v (mi/hr)
A	700	10	70
B	1600	120	13.3
C	2200	60	36.7
O	0	0	75

Solution

With the aid of graphical construction in Figure 11.3, the following can be calculated:

$U_{OB} = \frac{q_B - q_O}{k_B - k_O} = \frac{1600-0}{120-0} = 13.30$ mi/hr

$U_{AB} = \frac{q_B - q_A}{k_B - k_A} = \frac{1600-700}{120-10} = 8.18$ mi/hr

$U_{CB} = \frac{q_B - q_C}{k_B - k_C} = \frac{1600-2200}{120-60} = -10.00$ mi/hr

$\frac{be}{ae} = U_{OB} \rightarrow ae = \frac{be}{U_{OB}} = \frac{6.67}{13.3} = 0.5$ hr

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

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

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

$af = 0.64$ hr

So the impact of the truck lasts for 0.64 hour.

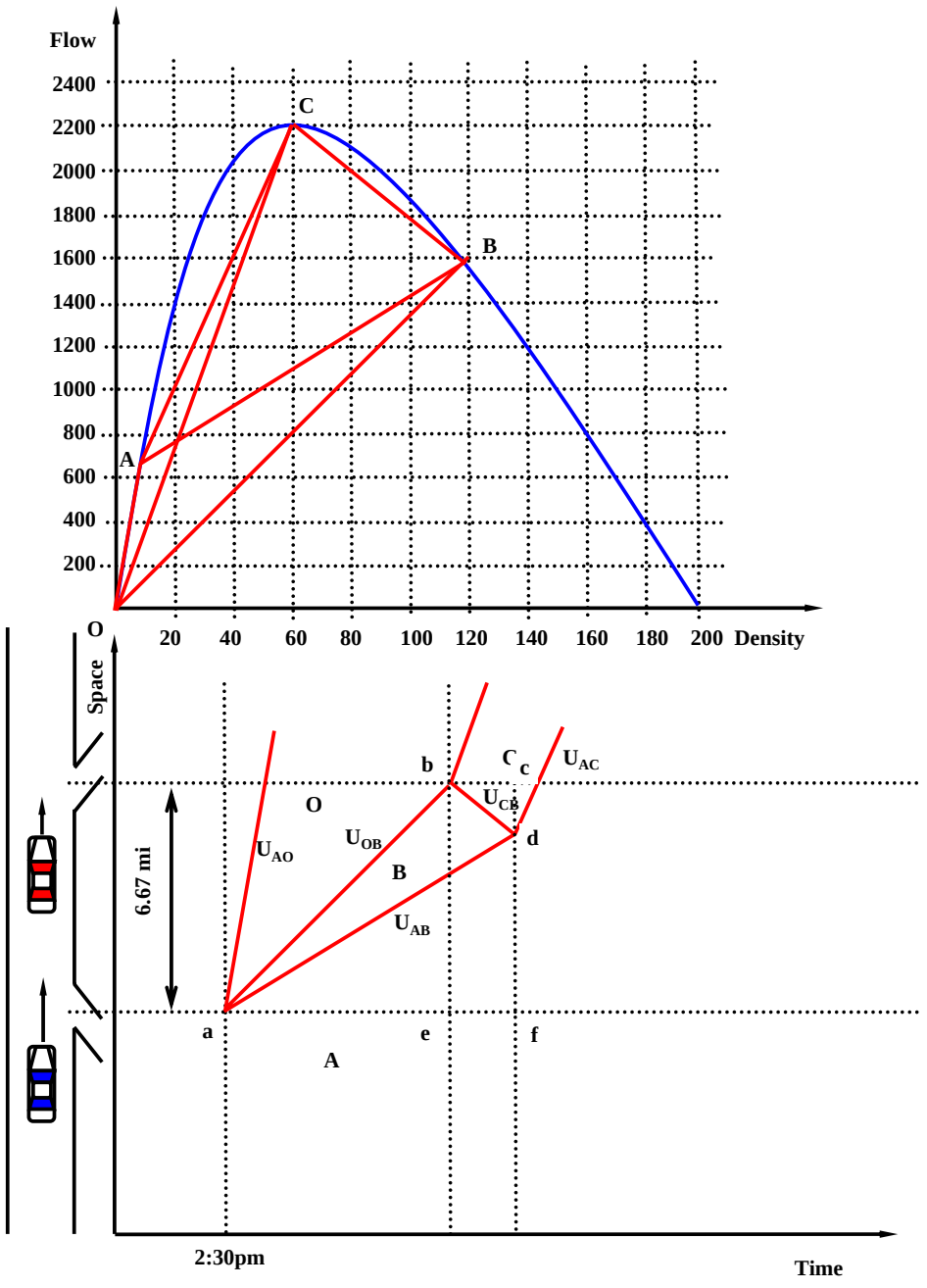


Figure 11.3: A moving bottleneck with constant demand

11.4 An ITS problem

On Wednesday 9:00 AM, there is an accident on northbound Interstate-91. The traffic operation center (TOC) has to decide how to cleanup the accident. After collecting information and communicate with highway patrol and emergency operator, the TOC determines that there are two alternatives:

1. Completely shut the Interstate off for 10 minutes, cleanup, and then reopen the Interstate for normal operation, or

2. Partially open the Interstate at reduced capacity, but the cleanup requires longer time - about 30 minutes - before normal operation can be resumed.

One of the concerns at the TOC is how long the queue will spill back because the queue on the Interstate will overflow via ramps and further block upstream surface streets. As a transportation engineering student, you are asked to offer you knowledge to help the TOC make decision.

More details in Table 11.4:

Table 11.4: Traffic data: an ITS problem

Condition	Description	q (veh/hr)	k (veh/mi)	v (mi/hr)
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

Find: which alternative creates a longer queue?

Solution

Alternative 1

With the aid of graphical construction in Figure 11.4, the following can be calculated:

$U_{AD} = \frac{q_D - q_A}{k_D - k_A} = \frac{0-2000}{200-40} = -12.5 \text{ mph} = \frac{x}{ac}$

$ac = \frac{x}{12.5}$

$U_{DC} = \frac{q_C - q_D}{k_C - k_D} = \frac{2200-0}{60-200} = -15.7 \text{ mph} = \frac{x}{bc}$

$bc = \frac{x}{15.7}$

$10 \text{ min} = \frac{1}{6} \text{ hr} = ac - bc = \frac{x}{12.5} - \frac{x}{15.7} = 0.016x$

$x = \frac{1}{6 \times 0.016} = 10.2 \text{ mi}$

Alternative 2

With the aid of graphical construction in Figure 11.5, the following can be calculated:

$U_{AB} = \frac{q_B - q_A}{k_B - k_A} = \frac{1100-2000}{145-40} = -8.57 \text{ mph} = \frac{x}{ac}$

$ac = \frac{x}{8.57}$

$U_{BC} = \frac{q_C - q_B}{k_C - k_B} = \frac{2200-1100}{60-145} = -12.94 \text{ mph} = \frac{x}{bc}$

$bc = \frac{x}{12.94}$

$30 \text{ min} = \frac{1}{2} \text{ hr} = ac - bc = \frac{x}{8.57} - \frac{x}{12.94} = 0.039x$

$x = \frac{1}{2 \times 0.039} = 12.7 \text{ mi}$

Comparing Alternative 1 in which the queue spills back 10.2 mi and Alternative 2 in which the queue spills back 12.7 mi, one concludes that Alternative 1 is better than 2 if queue length is of major concern.

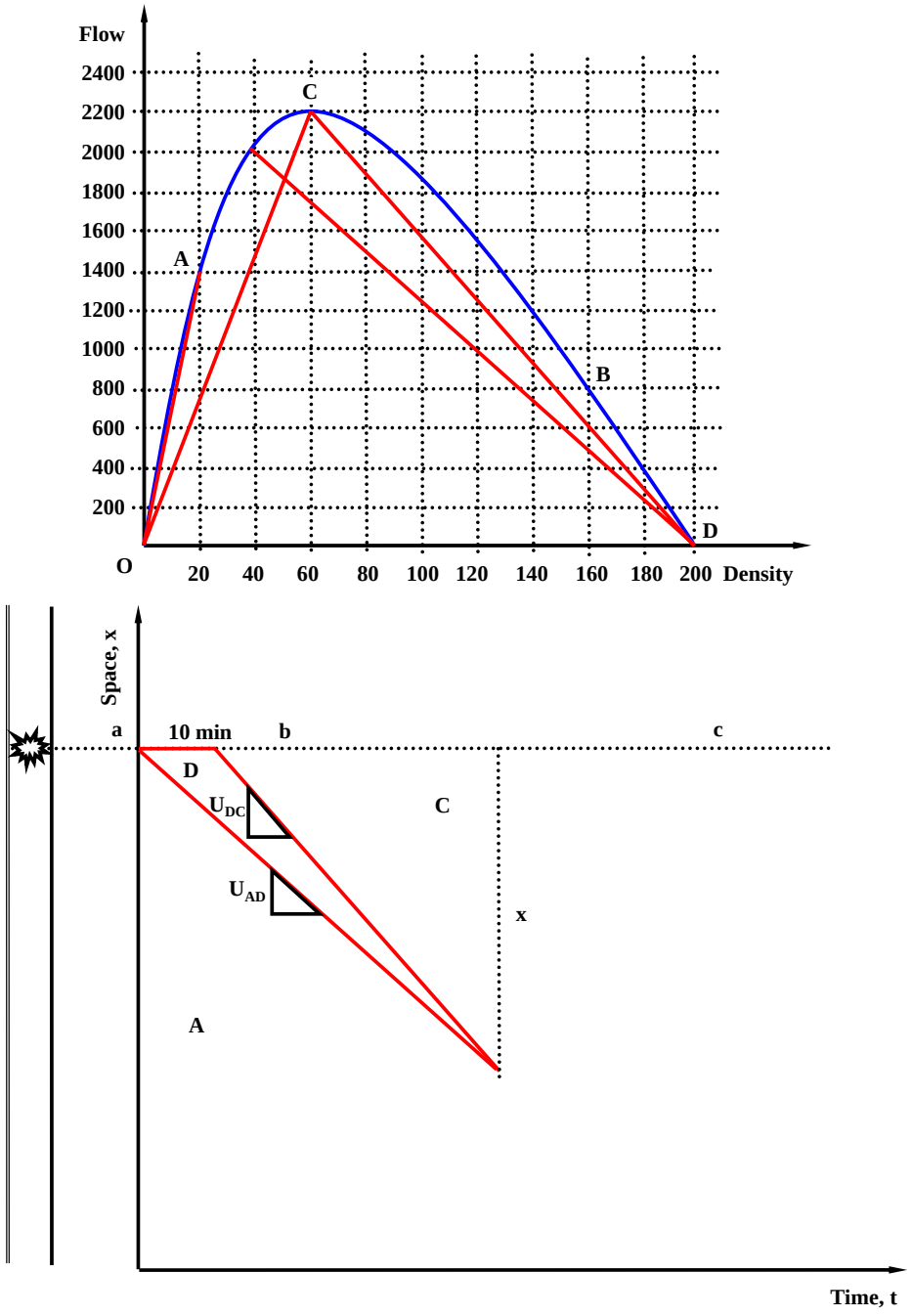


Figure 11.4: An ITS problem

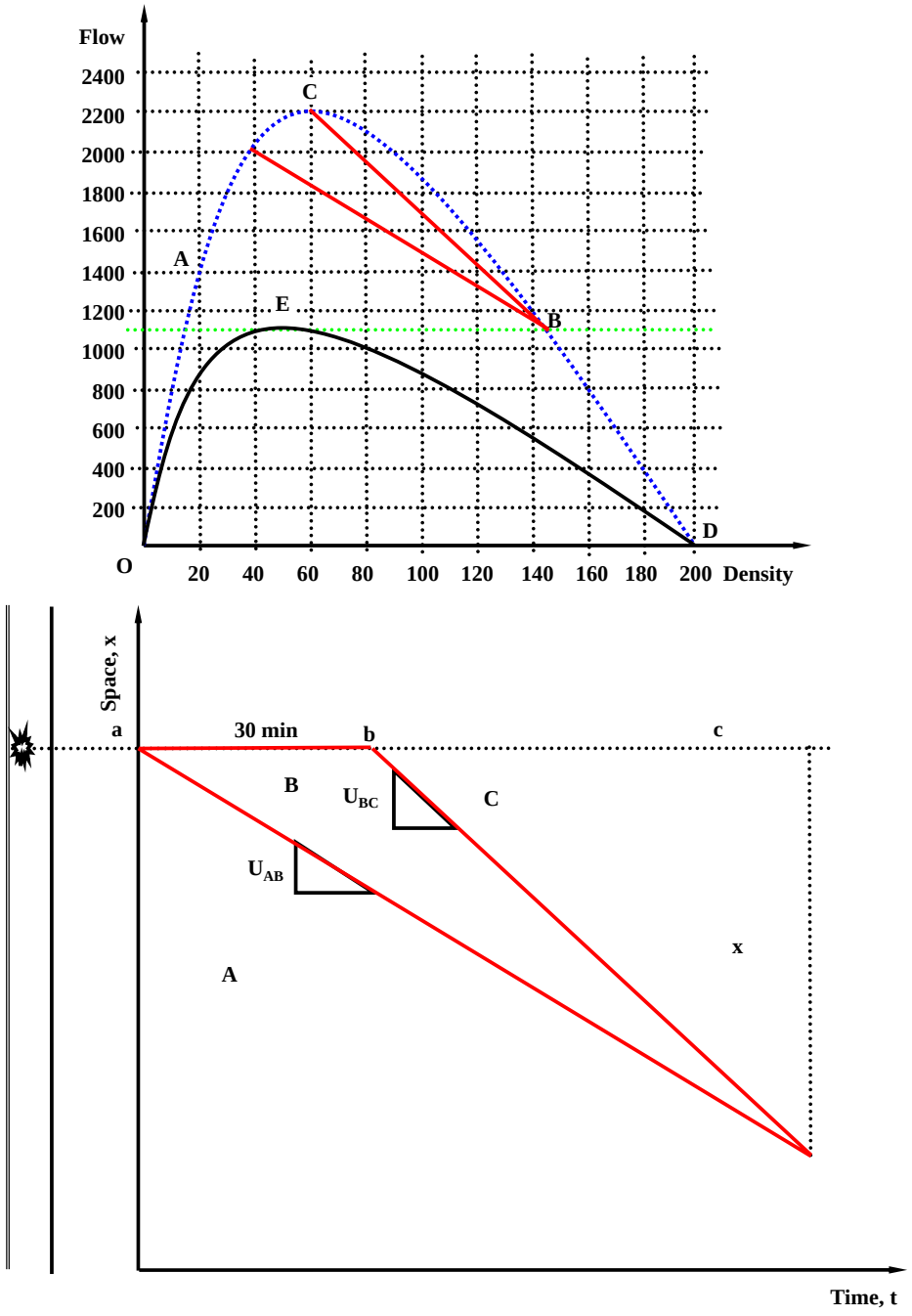


Figure 11.5: An ITS problem

Chapter 12

Numerical Solutions

Chapter 10 presented the LWR model and its solution procedure and Chapter 11 provided a few examples of LWR problems and their solutions. Note that these problems are solved graphically by manually working on a time-space diagram and using the method of characteristics. Though illustrative, the graphical approach it quite limited 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, initial and boundary conditions may be more complicated and time-varying. In these cases, the graphical approach is insufficient and sometimes infeasible. In addition, the purpose of solving an LWR model is frequently to enable traffic engineers the power to predict traffic dynamics so that they are able to expect where congestion might be and to develop management 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 (ITS) makes it possible to provide real-time traffic conditions and allow on-line prediction. Therefore, computerized solution to LWR model is essential to cope with more complicated real world problems, to provide real-time prediction, and to automate such prediction by developing on-line applications.

12.1 Discretization scheme

Computers are digital machines which can only work in a discrete fashion, so computerized solutions to LWR model have to be numerical and discrete. The first step to develop a computerized solution to discretize time and space. Figure 14.2 illustrates a time-space diagram where time t is the horizontal axis and space x is the vertical axis with a roadway drawn aside. The roadway is partitioned into a series of segments labeled as $j \in (0, 1, ...J)$. If x_0 is chosen as the reference point and segment length Δx is uniform, the location of the end of segment j is

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

.

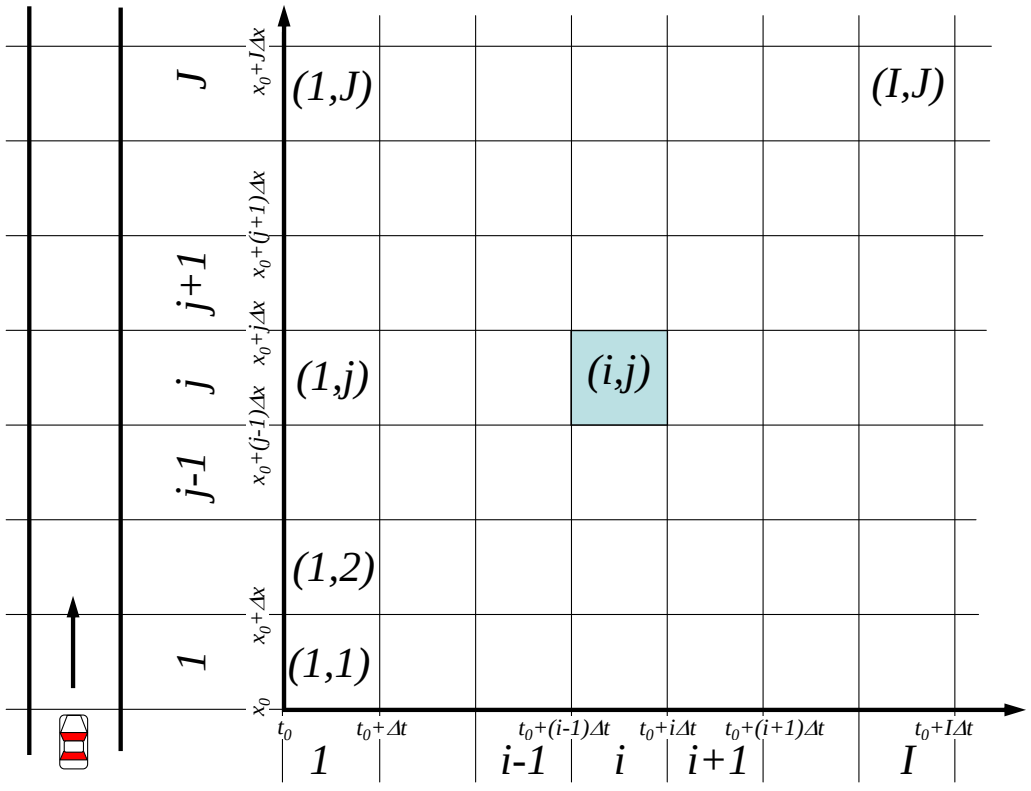


Figure 12.1: Discretization scheme

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

$$t_i = t_0 + \Delta t$$

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

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

where v_f is 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 LWR problem starts with initial conditions by determining number of vehicles contained each roadway segment one by one from upstream to downstream:

```
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 denote it as (i, j) and vehicles stored in segment j at the end of duration i as $n(t_i, x_j)$. The above listing can be re-written 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 inner:

```
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 are traversed and the solution is given as cell storage $[n(t_i, x_j)|i \in (1, 2, ..., I), j \in (1, 2, ..., J)]$ 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, a few numerical solutions to traffic dynamic problems are discussed in the following subsections.

12.2 FREFLO

FREFLO is an early (if no the earliest) computerized macroscopic traffic simulation model developed by Payne [115] in the late 1970s. Like LWR model, FREFLO consists of three equations with a discretization scheme shown in Figure 12.2. The first equation is conservation law:

```
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

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

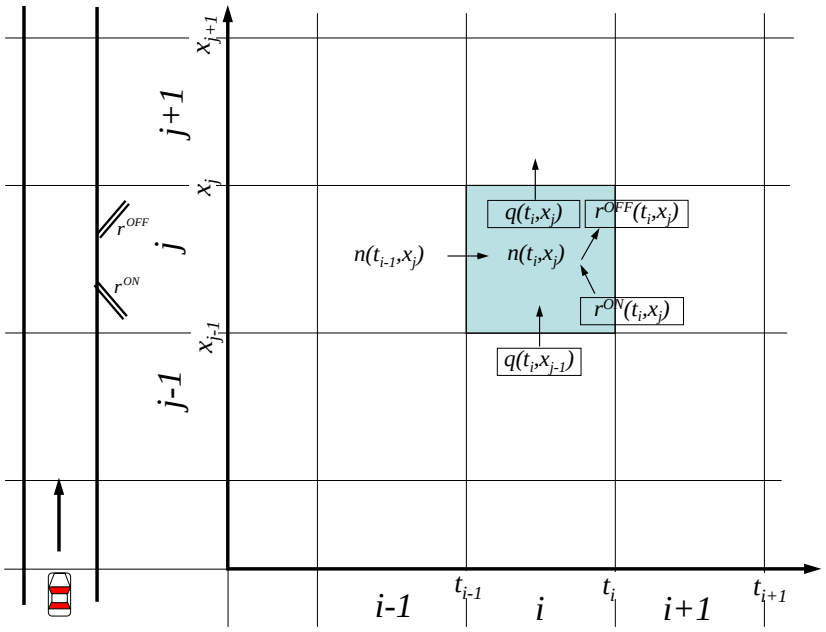


Figure 12.2: Discretization in FREFLO

where $g(t_i, x_j)$ is the net inflow via ramps, i.e. $g(t_i, x_j) = r^{ON}(t_i, x_j) - r^{OFF}(t_i, x_j)$. Note that $n = k\Delta x$, 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:

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

Differ from most first-order models which adopt a equilibrium speed-density relationship, FRELO 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

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

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

$$v = V(k) = 88.5(172 - 3.72k + 0.0346k^2 - 0.00119k^3)$$

which was an empirical speed-density relationship obtained by least-square 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.

12.3 FREQ

FREQ is a computerized macroscopic traffic simulation model developed by May et al in the early 1980s [87]. Its underlying algorithm is not publicly available.

12.4 KRONOS

KRONOS is another computerized macroscopic traffic simulation model developed by Michalopoulos [91] in the middle 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

```
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 an illustration in Figure 12.3. Mathematically, this is equivalent to:

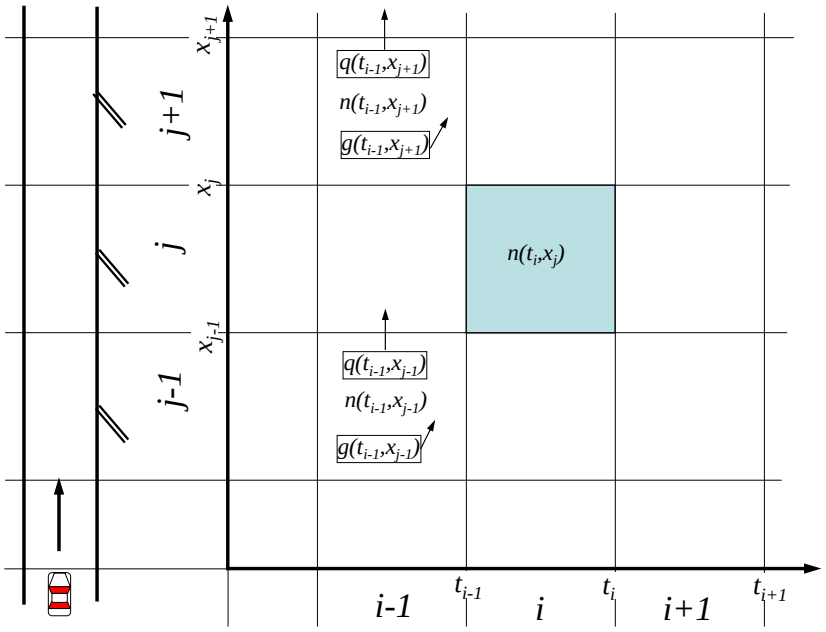


Figure 12.3: Discretization in KRONOS

$$\begin{aligned} n(t_i, x_j) &= \frac{n(t_{i-1}, x_{j+1}) + n(t_{i-1}, x_{j-1})}{2} \\ &\quad - \frac{\Delta tq(t_{i-1}, x_{j+1}) - \Delta tq(t_{i-1}, x_{j-1})}{2} \\ &\quad + \frac{\Delta tg(t_{i-1}, x_{j+1}) + \Delta tg(t_{i-1}, x_{j-1})}{2} \end{aligned}$$

Note that $n = \Delta xk$, the above equation becomes

$$\begin{aligned}\Delta xk(t_i, x_j) &= \frac{\Delta xk(t_{i-1}, x_{j+1}) + \Delta xk(t_{i-1}, x_{j-1})}{2} \\ &\quad - \frac{\Delta tq(t_{i-1}, x_{j+1}) - \Delta tq(t_{i-1}, x_{j-1})}{2} \\ &\quad + \frac{\Delta tg(t_{i-1}, x_{j+1}) + \Delta tg(t_{i-1}, x_{j-1})}{2}\end{aligned}$$

Hence,

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

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 Greenshields model [50]

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

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

The boundary condition is given as: $q(t_i, x_0)$, $i = 0, 1, 2..., I$ and $g(t_i, x_j)$, $i = 0, 1, 2..., I$ and $j = 0, 1, 2..., J$

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

Chapter 13

Simplified Theory of K-Waves

The simplified theory of kinematic waves (K-Waves) was formulated by Newell [100, 101, 102] in the early 1990s. The theory was published on a trio of papers with the first paper addressing general theory, the second being devoted to queuing at a freeway bottleneck, and the third multi-destination flows.

13.1 Triangular flow-density relationship

K-Waves was proposed as a (graphical) solution to LWR model under a special condition: the underlying flow-density relationship is a triangular one with jam density K and capacity Q, see Figure 13.1.

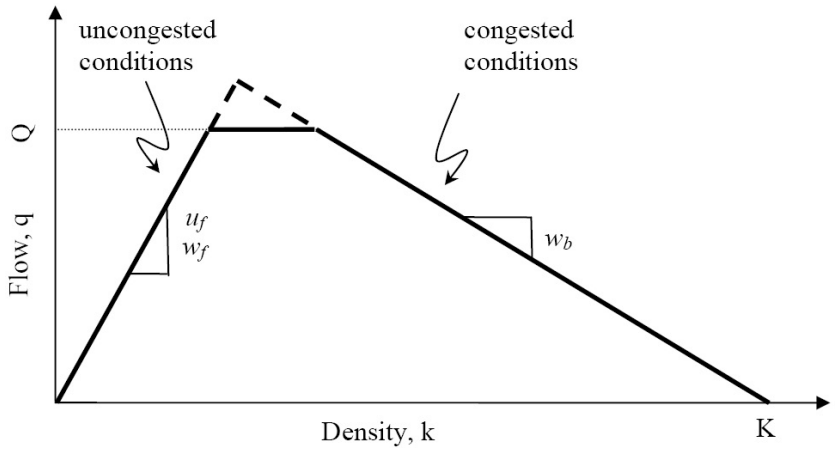


Figure 13.1: Triangular flow-density realtionship

Referring back to Chapter 10 Figure 10.5, a point on the flow-density curve uniquely defines the operating condition of a stream of traffic. The speed of 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 free-flow speed v_f. As a special property of the triangular flow-density relationship, v_f applies to all uncongested conditions.

13.2 Forward wave propagation

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

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 $T = \frac{x_j - x_{j-1}}{v_f}$ assuming that vehicles preserve their order. The traffic also carries a kinematic wave whose speed w_f happens to be v_f, as noted above, so it is equivalent to say 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 13.2 where the profile $N(t, x_j)$ is simply a horizontal transportation of profile $N(t, x_{j-1})$ to the right by T:

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

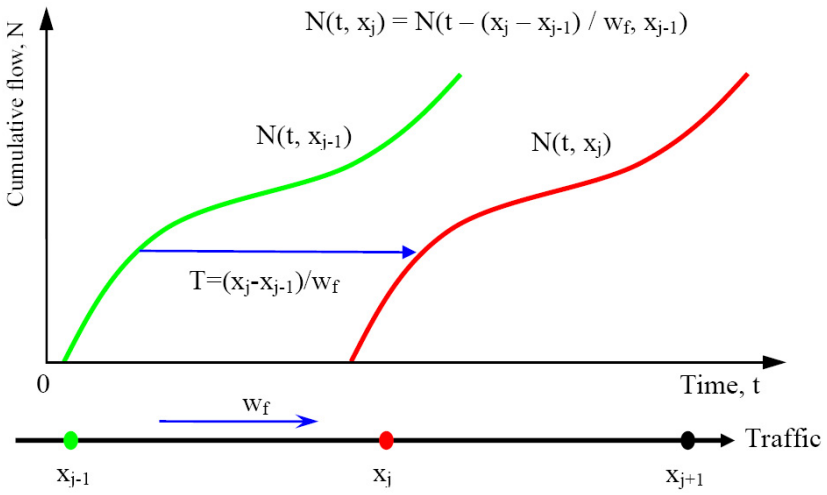


Figure 13.2: Forward wave propagation

13.3 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 13.3. Then the kinematic wave carried by the traffic will propagate backward at speed w_b. Hence, 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 $n = K_j(x_{j+1} - x_j)$:

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

13.4 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 exceeding the capacity, this is equivalent to say 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. Run 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

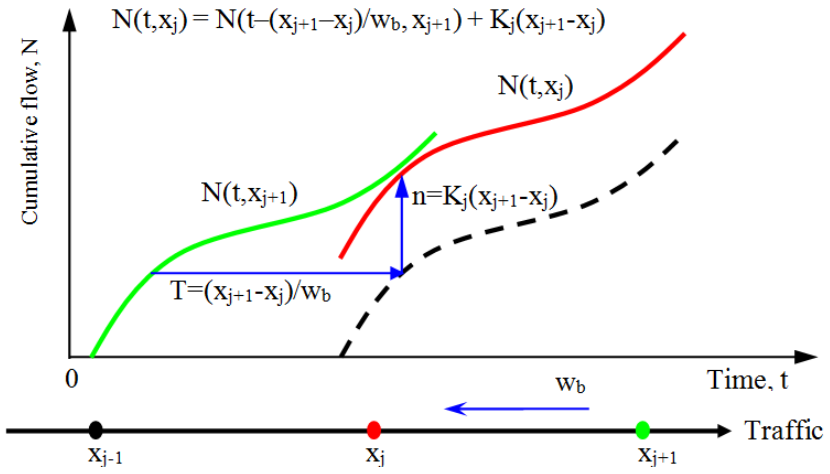


Figure 13.3: Backward wave propagation

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 13.4.

13.5 Minimum principle

Intuitively, the minimum principle means that any point on a roadway x_j cannot admit more vehicles than what arrive from upstream $N^{up}(t, x_j)$, what is allowed by local capacity $N^Q(t, x_j)$, and what the downstream is able to receive $N^{dn}(t, x_j)$. Graphically, this is to superimpose the above three 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 13.5:

$$N(t, x_j) = min[N^{up}(t, x_j), N^Q(t, x_j), N^{dn}(t, x_j)]$$

13.6 Single bottleneck

In Figure 13.5, 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 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:

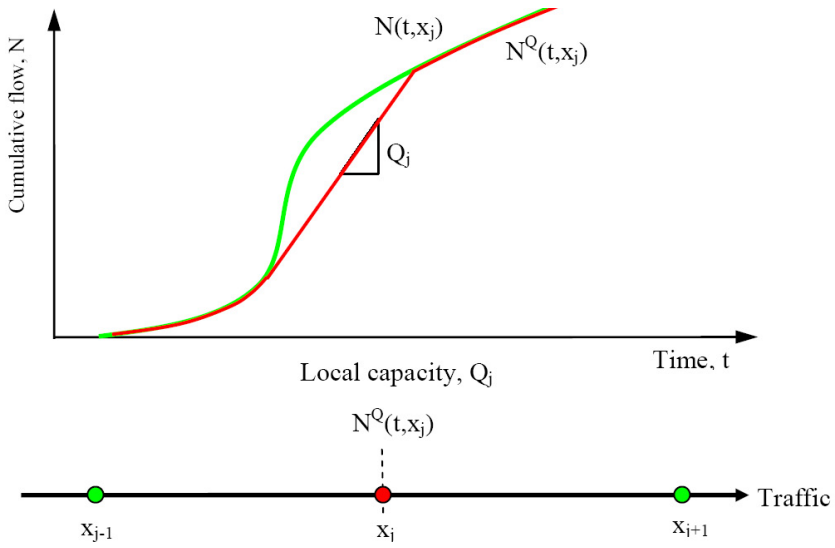


Figure 13.4: Flow constrained by local capacity

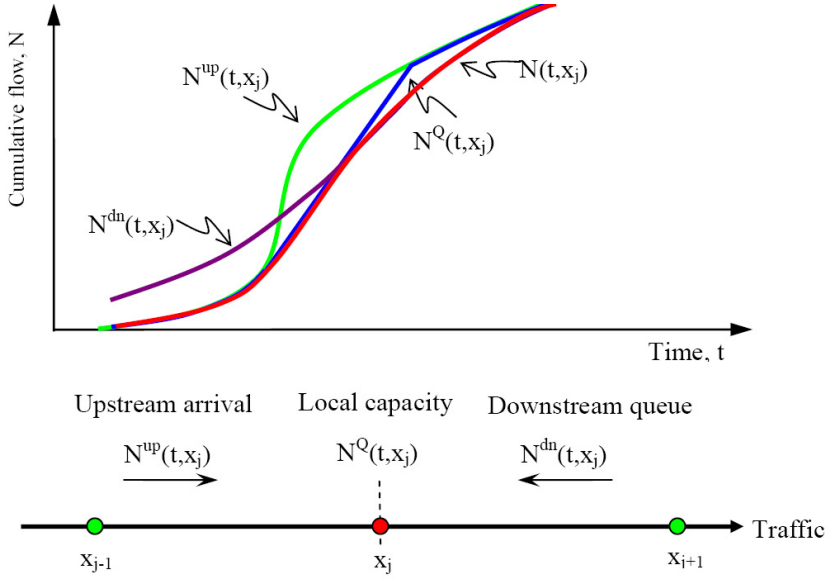


Figure 13.5: The minimum principle

- 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 queue at time t.

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

- 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 13.2), cumulative flow arriving at location slightly upstream of the bottleneck (to the left of x_j), x_j^- is:

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

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

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

From wave backward propagation (Figure 13.3), the cumulative flow allowed to depart is:

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

Considering local capacity (Figure 13.4), cumulative flow departing x_j^+ should not exceed $N_Q(t, x_j^+)$.

Therefore, based on the minimum principle (Figure 13.5), the cumulative flow actually departed at x_j^+ is

$$D(t, x_j^+) = min[N^{up}(t, x_j^+), N_Q(t, x_j^+), N_{dn}(t, x_j^+)]$$

Assume on-ramp traffic, $A_j(t)$, has priority over mainline traffic and can always bypass any queue at the bottleneck (this is a limitation of the simplified K-Waves theory since there is no queueing at 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 13.6.

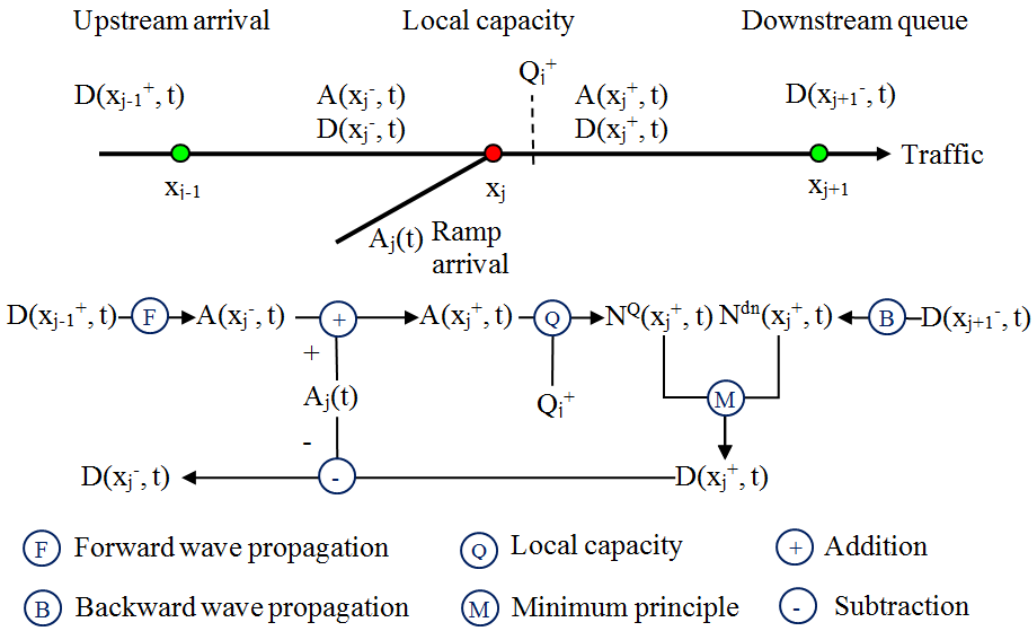


Figure 13.6: Single bottleneck

13.7 Computational algorithm

With the above preparation, traffic flow on a freeway involve multiple segments and bottlenecks can be numerically modeled as follows. First, the time and space is partitioned using the discretization scheme in Chapter 12, resulting a lattice shown in Figure 13.7. Next, starting from initial conditions, the numerical solution procedure outlined in Chapter 12 is applied. At each lattice point (t_i, x_j), the cumulative arrival and departure flows are determined as follows:

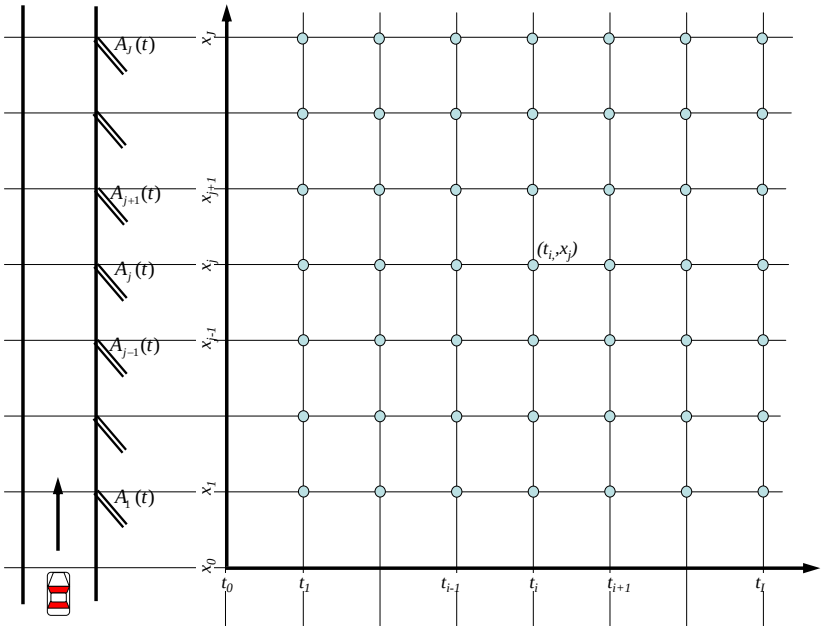


Figure 13.7: K-Waves lattice

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^+

$$D(t_i, x_j^+) = min[N^{up}(t_i, x_j^+), N^Q(t_i, x_j^+), N^{dn}(t_i, x_j^+)]$$

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:

$$\begin{array}{c} A(t_1, x_1^-),\ D(t_1, x_1^-),\ A(t_1, x_1^+),\ D(t_1, x_1^+) \\ \cdots \\ A(t_1, x_J^-),\ D(t_1, x_J^-),\ A(t_1, x_J^+),\ D(t_1, x_J^+) \\ A(t_2, x_1^-),\ D(t_2, x_1^-),\ A(t_2, x_1^+),\ D(t_2, x_1^+) \\ \cdots \\ A(t_I, x_J^-),\ D(t_I, x_J^-),\ A(t_I, x_J^+),\ D(t_I, x_J^+) \end{array}$$

13.8 Further notes on K-Waves theory

The above discussion summarizes the first two papers of Newell's simplified K-Waves theory [100], [101] involving bottlenecks with on-ramps only.

The third paper [102] incorporates off-ramps into consideration and, hence, multiple destination flows. Discussion on this subject is quite involved and readers are encouraged to read the original paper for full information. In addition, supplementary information on simplified K-Waves can be found in Son and Hurdle [138, 65] for model validation and extracting traffic dynamics information and Ni and Leonard [104, 110] for 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. Such a partitioning is only necessary 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 cell-based models such as FREFLOW, KRONOS, and Cell Transmission Model whose accuracy relies on cell size (length of segment). Consequently, the requirement of computation and storage can be significantly reduced.

In addition to the assumption of a triangular flow-density relationship, another limitation of the K-Waves theory is its assumption that on-ramp traffic has priority over mainline traffic and can always bypass any queue at a bottleneck. Consequently, the K-Waves theory is unable to model network traffic where queueing at ramps has to be accounted for. Further attempt to address this issue can be found in Ni and Leonard [104, 110] where on- and off-ramp queuing models have been proposed, based on which the K-Waves theory was extended to network flows.

Chapter 14

Cell Transmission Model

The cell transmission model (CTM) was proposed by Daganzo [18, 21] in the middle 1990s. The model was presented in two papers with the first addressing mainline traffic and the second network traffic.

14.1 Minimum principle

Figure 14.1 is a reproduction of the triangular flow-density relationship presented in Figure 13.1. 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.

A vertical line at any density k will intersect the three sections at height kw_f, q_m, and $(K-k)w_f$. 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 upstream arrival rate, local capacity, and the rate allowed by downstream conditions.

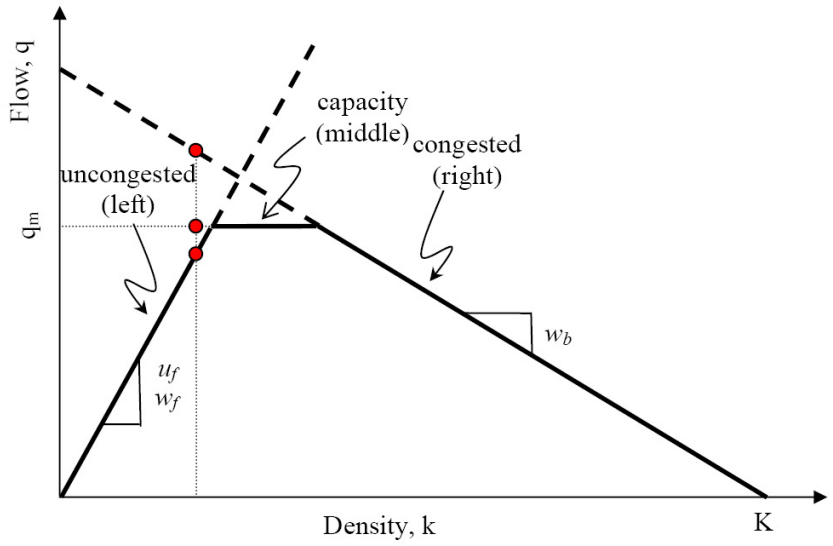


Figure 14.1: Triangular flow-density realtionship

14.2 Mainline scenario

CTM uses the same discretization scheme presented in Chapter 12 and the scheme is reproduced in Figure 14.2 for convenience. 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 amount of vehicles that can move into segment j, $y_j(t_i)$, is found by multiply both sides by Δt:

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

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

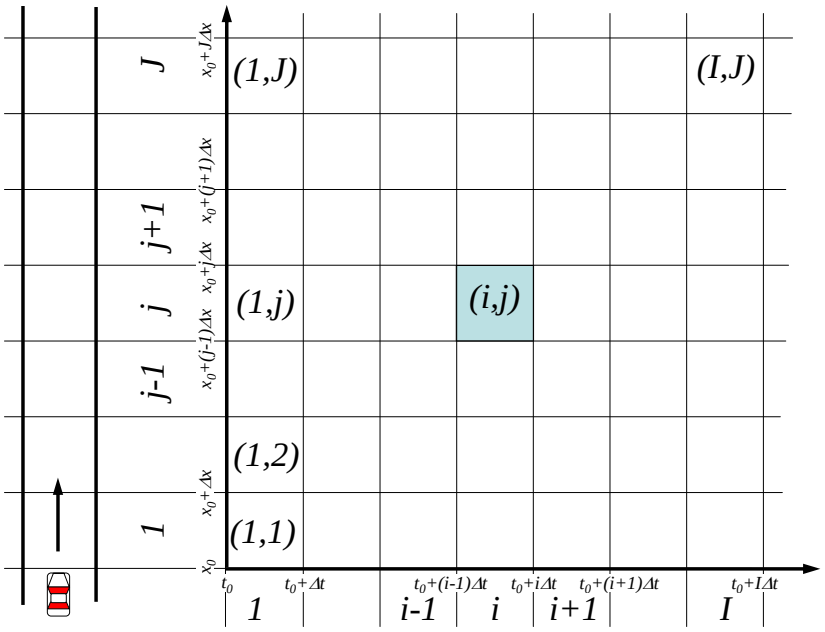


Figure 14.2: Discretization scheme

form:

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

i.e.,

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

The above equation stipulates that the amount of vehicles that can move into segment j, $y_j(t_i)$, is constrained by the following:

- 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

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

where $S_{j-1} = min[n_{j-1}(t_{i-1}), q_m \Delta t]$ represents flow being sent from upstream 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 by 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

$$n_j(t_i) = n_j(t_{i-1}) + y_j(t_{i-1}) - y_{j+1}(t_{i-1})$$

14.3 Merge scenario

To be able to address network traffic, a queuing model is needed for a merge where two streams of traffic flow into one. The merge 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 14.3. 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 $(j-1)$' actually sends y'_{j-1} 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 merge 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 14.4 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. Emit a line from the origin O with slope $\frac{p'_{j-1}}{p_{j-1}}$ and the

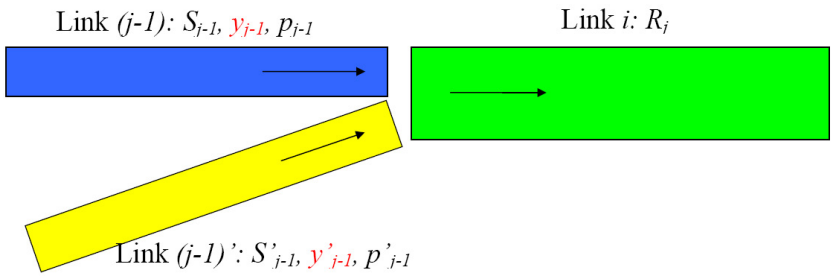


Figure 14.3: A freeway merge

line intersect the rectangle at point C. Curve ACO denotes the collection of solutions and reason is the following.

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

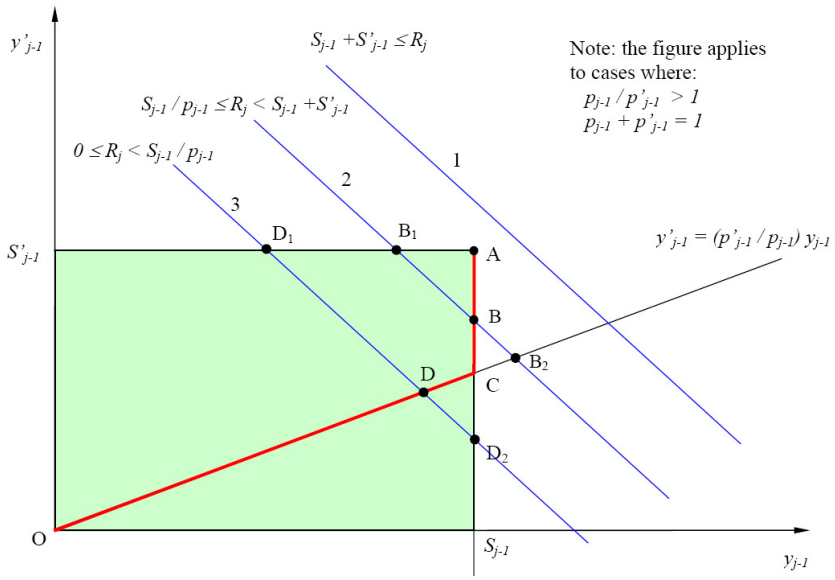


Figure 14.4: Queuing at a freeway merge

1. Supply exceeds demand

This is to say that $S_{j-1} + S'_{j-1} \le R_j$. Physically, this means that link j is able to receive more vehicles than the total to be sent from both upstream links.

For example, $S_{j-1} = 100$, $S'_{j-1} = 80$, $R_j = 200$. 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 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:

$$\begin{cases} y_{j-1} = S_{j-1} \\ y'_{j-1} = S'_{j-1} \end{cases} \quad \text{if } R_j \geq S_{j-1} + S'_{j-1} \tag{14.1}$$

This solution corresponds to vertex A in the figure.

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, $S_{j-1} = 100$, $S'_{j-1} = 80$, $R_j = 160$. The priority rules stipulates 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 $160 \times \frac{1}{4} = 40$ vehicles. Since the mainline only has 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. Since 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. Using 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 out 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 out 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 out of the 80 vehicles from the on-ramp. Since link $j-1$ can not depart more vehicles than it has, this solution is incorrect.

Based on 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:

$$\begin{cases} y_{j-1} = mid[S_{j-1}, R_j - S'_{j-1}, p_{j-1}R_j] \\ y'_{j-1} = mid[S'_{j-1}, R_j - S_{j-1}, p'_{j-1}R_j] \end{cases} \quad \text{if } R_j < S_{j-1} + S'_{j-1} \tag{14.2}$$

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. Still use the above example except that $R_j = 120$. The priority rules stipulates that link $j-1$ can send as a maximum $120 \times \frac{3}{4} = 90$ vehicles to link j and the on-ramp $120 \times \frac{1}{4} = 30$ vehicles. Since both upstream links have more vehicles than they are able to send, the priority rule takes control, i.e. 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 the priority line between points C and O. Again, Line 3 consists of all points whose coordinates sum up to R_j. Using 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 the following:

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 out 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 out 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 is used to admit 20 out 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 merge model is:

$$\begin{cases} \begin{cases} y_{j-1} = S_{j-1} \\ y'_{j-1} = S'_{j-1} \end{cases} & \text{if } R_j \geq S_{j-1} + S'_{j-1} \\ \begin{cases} y_{j-1} = mid[S_{j-1}, R_j - S'_{j-1}, p_{j-1}R_j] \\ y'_{j-1} = mid[S'_{j-1}, R_j - S_{j-1}, p'_{j-1}R_j] \end{cases} & \text{if } R_j < S_{j-1} + S'_{j-1} \end{cases} \tag{14.3}$$

14.4 Diverge scenario

A queuing model is also needed for a diverge where one stream of traffic splits into two. The diverge consists of one upstream link $j-1$ and two downstream links (e.g. a mainline link j and an off-ramp j'), see Figure 14.5. 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 pre-determined: 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 first-in-first-out queuing discipline and, if a vehicles fails to depart, it holds up all vehicles behind. The question here is to determine actual outflow of

link $j-1$, y_{j-1}, among which how many are destined for link j, y_j and how many for the off-ramp, y'_j.

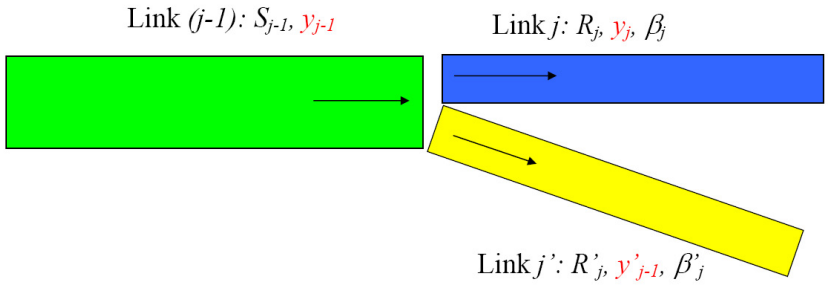


Figure 14.5: A freeway diverge

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

$$\begin{cases} y_{j-1} = y_j + y'_j \leq S_{j-1} \\ y_j = \beta y_{j-1} \leq R_j \\ y'_j = \beta' y_{j-1} \leq R'_j \end{cases} \tag{14.4}$$

Hence,

$$y_{j-1} = min[S_{j-1}, \frac{R_j}{\beta}, \frac{R'_j}{\beta'}]$$

Consequently, one obtains $y_j = \beta y_{j-1}$ and $y'_j = \beta' y_{j-1}$.

Chapter 15

High-Order Models

Macroscopic traffic flow models discussed so far, including both analytical and numerical models, have been focused on LWR model [80, 130] and its variants. At the center of these models is mass or vehicle conservation which can be mathematically expresses as a first-order partial differential equation (PDE):

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

where k an 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.

15.1 High-order models

In essence, conservation law takes several forms and mass or vehicle conservation is one (and perhaps the simplest) of them. Other forms of conservation, for example, can be conservation of linear momentum and conservation of energy which involve high-order PDEs. If a model involves such equations, it is classified as a high-order model. Listed below are a few examples of such models.

PW model (1971)

Proposed by Payne [114] and independently by Whitham [156], the PW model consists of a system of two equations, the first of which is the conservation of mass given in LWR model. 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 12 is the numerical solution based ont PW model.

Phillips' model (1979)

Based on kinetic theory, Phillips [118] developed a model which incorporated 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 as defined above.

Kühne's model (1984)

Kühne [71, 72] 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 where c_0 is sound speed, and η is a viscosity constant.

Kerner and Konhäuser's model (1993)

Kerner and Konhäuser [70] 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:

$$\begin{cases} \frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0 \\ \frac{\partial v}{\partial t} = \lambda(V_e(k) - v) - \frac{c_0^2}{k}\frac{\partial k}{\partial x} + \frac{1}{k}\frac{\eta \partial v}{\partial x} \end{cases}$$

Michalopoulos' model (1993)

Michalopoulos, et al [92] 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 mode is more accurate and computationally more efficient.

$$\begin{cases} \frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0 \\ \frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = \frac{1}{\tau}(v_f - v) - G\frac{\partial v}{\partial t} - \nu k^{\beta}\frac{\partial k}{\partial x} \end{cases}$$

where $G = \mu k^{\varepsilon} g$, and μ, ν, ε, and β are all constant parameters, and v_f is free-flow speed.

Zhang's model (1998)

Zhang [163] proposed a non-equilibrium traffic flow model which is based on both empirical evidence of traffic flow behavior and basic assumptions on drivers' reaction 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.

$$\begin{cases} \frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0 \\ \frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = \lambda(V_e(k) - v) - k(V_e'(k))^2\frac{\partial k}{\partial x} \end{cases}$$

Treiber's model (1999)

Treiber, et al. [144] derived macroscopic traffic equations from specific gas-kinetic equations and the resulting partial differential equations for vehicle density and average speed contain a non-local interaction term which is very favorable for a fast and robust numerical integration, so that several thousand freeway kilometers can be simulated in real time.

$$\begin{cases} \frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0 \\ \frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = \lambda(V_e(k) - v) + \frac{1}{k}\frac{\partial kAv^2}{\partial x} - \frac{V_e A(k)}{\tau A(k_j)}[\frac{k_\alpha T v}{1-\frac{k_\alpha}{k_j}}]^2 B(\delta_v) \end{cases}$$

where $A = A(k)$ is a density-dependent function, k_α is the density at point x_α ahead of x, $B(\delta_v)$ is a macroscopic interaction term, and $V_e(k)$ is the normal equilibrium speed-density relationship.

15.2 Relating continuum flow models

Starting from mass or vehicle conservation, a variety of continuum flow models have been developed by supplying additional assumptions. Generally, these models can be summarized by the following model [160]:

$$\begin{cases} \frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = g(t,x) \\ \frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = \frac{1}{\tau}(V_e(k) - v) + \frac{1}{k}\frac{\partial P}{\partial x} \end{cases}$$

where $V_e(k,v)$ is the generalized equilibrium speed-density relationship. $P(k,v)$ is the traffic pressure, and τ is the relaxation time, which is the time constant of the regulating traffic speed v to the equilibrium speed V_e. $g(t,x)$ is net ramp inflow. For a highway without on- 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 applying different traffic pressure P, relaxation time τ, and the generalized equilibrium speed U_e. For example,

- if $\tau = 0$ and $P = 0$, LWR model is resulted;
- if $P = -\frac{V_e(k)}{2\tau}$ with $V_e(k,v) = V_e(k)$, PW model is resulted;
- if $P = k\Theta$ with $\Theta = \Theta_0(1 - \frac{k}{k_j})$ and k_j the jam density, Phillips' model is resulted;

- if $P = k\Theta_0 - \eta\frac{\partial v}{\partial x}$, Kerner and Konhäuser model is resulted;
- Michalopoulos' model is resulted if $P = \frac{\nu}{\beta+2}k^{\beta+2}$, where ν is an anticipation parameter, β a dimensionless constant, and $V_e(k) = v_f$ where v_f is free-flow speed;
- Zhang's model is resulted if $P = \frac{1}{3}k^3V_e'^2(k)$ with $V_e'(k) = \frac{dV_e(k)}{dk}$.
- Treiber's model is resulted if $P = Akv^2$ where $A = A(k)$ is a density-dependent function, and $V_e(k, v) = V_e(k)\{1 - \frac{A}{A(k_j)}[\frac{k_\alpha Tv}{1-\frac{k_\alpha}{k_j}}]^2 B(\delta_v)\}$.

15.3 Relative merits of continuum models

Daganzo [20] noted that, as a first-order continuum flow model, LWR model is proposed on dense traffic with equilibrium and it is flawed for light traffic. This is because, when passing is allowed, 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 LWR model. When passing is restricted, LWR model produces unsatisfactory results in the following three aspects. First, LWR model predicts abrupt speed change when a vehicle passes through a shockwave, an action that is unrealistic in real world. Second, LWR model fails to predict instabilities of top-start traffic. Third, LWR model assumes zero reaction time which does not happen in real world. Readers are referred to Daganzo's original paper for full information.

Given these deficiencies, continuum flow models developed so far have been trying to fix the deficiencies and almost all these models follow the direction of incorporating a momentum conservation equation. An early attempt to fix the deficiencies in LWR model was made by Prigogine [124] who proposed a kinetic model incorporating a speed distribution to address platoon dispersion. A decade later, Payne [114] and Whitham [156] 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 shockwave. This seminal work has inspired many thoughts in analytical explanation of shockwave behavior and, thus, has spawned several variants, among which are Phillips [118], Kühne [71, 72], Kerner and Konhäuser [70], Michalopoulos, et al [92], Zhang [163], Treiber, et al. [144], etc.

Several deficiencies are found in PW model [20]. First, it does not remove all the shock waves. Second, as reported by de1 Castillo, et al. [25], vehicles in PW model can adjust their speeds in response to disturbance from behind, while in reality vehicles typically respond to their leaders. Third, PW model incorporates a momentum equation which is derived from a car-following model. This momentum equation neglects second- and higher-order terms of spacings and speeds which may not be negligible when spacings and speeds are not slowly varying. Fourth, PM model as well as other high-order model always produce wave speeds that are greater than traffic speeds. This is an unattractive property to 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 [90, 78, 20] tend to believe that high-order models, despite their added complexity and additional parameters, do not improve LWR model .

15.4 Taxonomy of macroscopic models

Figure 15.1 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 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 pane. These models include KRONOS, K-Waves (though this is a graphical solution involving discrete space but continuous time), and CTM. K-Waves model was further extended to network traffic by Ni and Leonard [104, 110].

Central to high-order models are equations of mass and momentum con-

servation. These models include Payne and Whitham [114, 156], Phillips [118], Khne [71, 72], Kerner and Konhuser [70], Michalopoulos [92], Zhang [163], Treiber [144], etc. FREFLO is a numerical model derived from Payne and Whitham [114, 156].

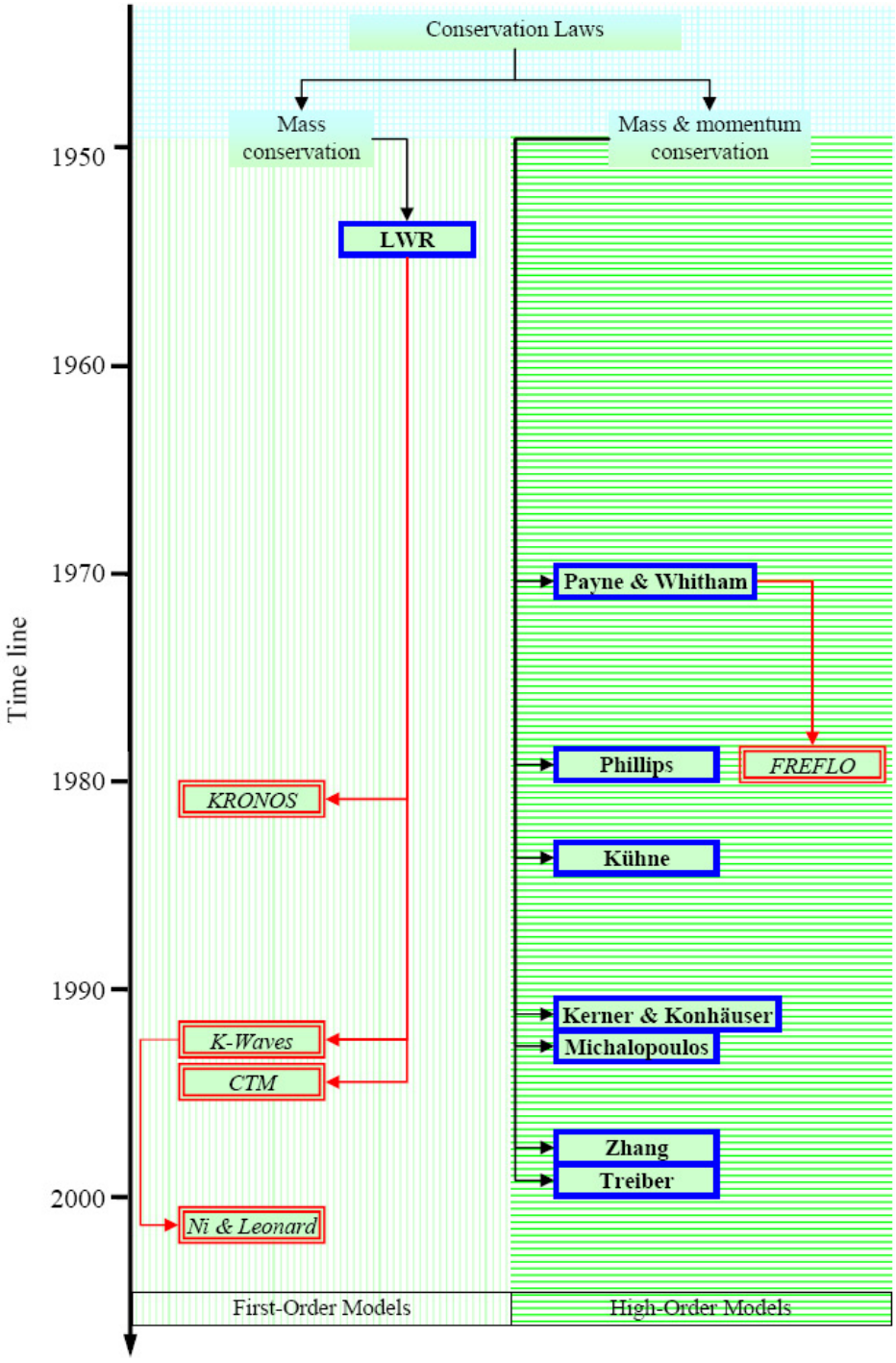


Figure 15.1: Taxonomy of macroscopic models

Part III

Microscopic Modeling

Chapter 16

Microscopic Modeling

Models presented in Chapters 5 through Chapter 15 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 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.

16.1 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 16.1 illustrates the decision-making scenario faced by a driver who is about to travel from UMass Amherst (Point C) to Boston (Point D). The driver has at least three options:

- I-90 (Masspike). The bottom route which is the fastest route if there

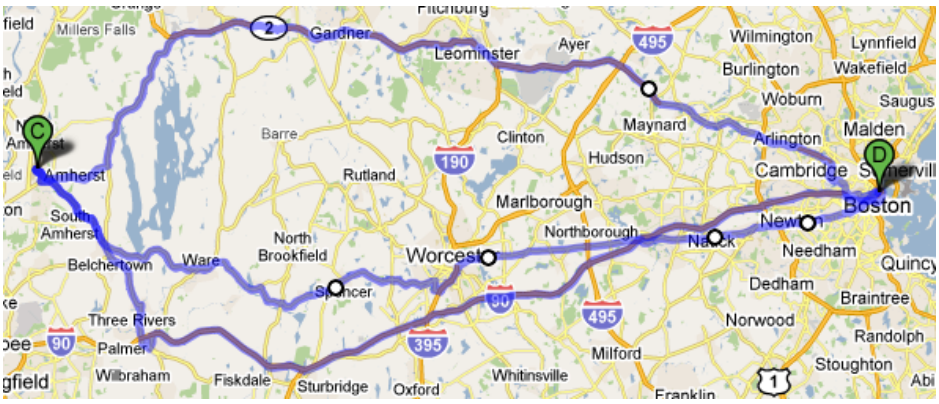


Figure 16.1: Making decision at the strategic level

is no congestion, and the toll is about 2 to 3 dollars.

- Rt. 2. The top route which is a scenic, rural highway that is rarely congested.
- Rt. 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 ahead. A 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 out of a set of options based on some utilities and constraints.

After the driver has chosen a route (e.g., Masspike) and is traveling down the road, a *tactical* decision will come up sooner or later that involves a medium geographical scope and a medium time frame. For example, Figure 16.2 illustrates that the driver need 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 of about a few km in geographical scope and a few minutes in time frame. Again, a lane-changing model is typically a discrete choice model that determines a target lane out of 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 16.3 illustrates

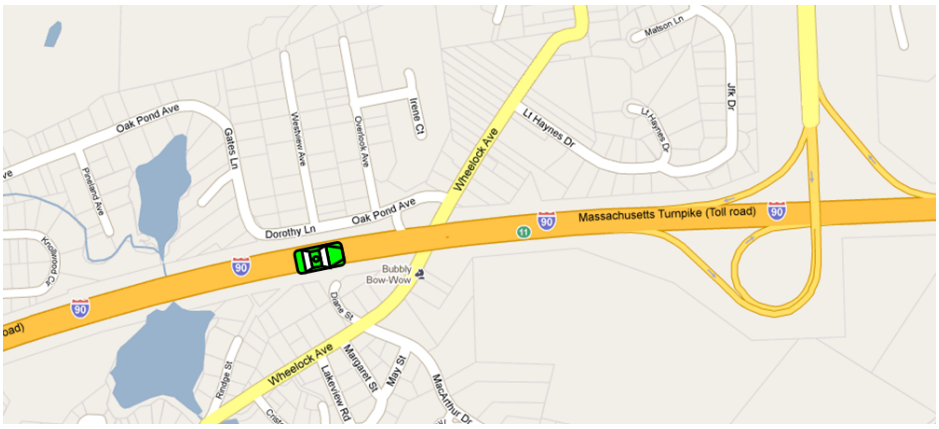


Figure 16.2: Making decision at the tactical level

Figure 16.3: Making decision at the operational level

that the driver (the circled one) is following another vehicle in a context of about tens of meters in geographical scope and a few seconds in time frame. The driver needs to make a **car following** decide on how 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 stressful 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 out for gaps in the adjacent lane and switching to that lane when an acceptable gap becomes available.

Therefore, based on the geographical scope and time frame involved, microscopic models can fall into the following three broad categories:

- At the strategic level: route-choice models
- At the tactical level: lane-changing models
- At the operational level: car-following and gap-acceptance models

16.2 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 formal discussion on car-following models, it helps to summarize the notation to be used. Figure 18.1 illustrates a few vehicles traveling on a one-lane highway. These vehicles $(1, 2, ..., i-1, i, i+1, ..., I)$ are numbered cumulatively with lower-numbered vehicles in front, e.g., vehicle #1 leads vehicle #2. The locations or displacements of vehicles are measured from a common but arbitrary reference point.

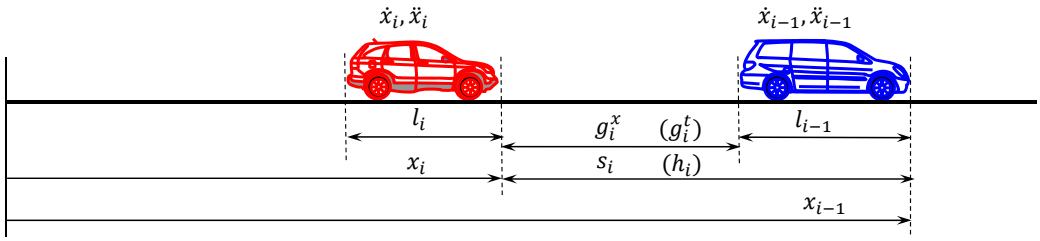


Figure 16.4: A car following scenario

i	vehicle ID, $i = 1, 2, ..., I$.
$x_i(t)$	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 wherever 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 its leading vehicle $i-1$ at time t.
$g_i^x(t)$	the distance gap between vehicle i and its leading vehicle $i-1$ at time t.
$h_i(t)$	the headway between vehicle i and its leading vehicle $i-1$ at time t.
$g_i^t(t)$	the time gap between vehicle i and its leading vehicle $i-1$ at time t.

16.3 Benchmarking Scenarios

The upcoming chapters will introduce some microscopic car-following models. These models were formulated using a variety of modeling philosophies and appeared in different times. It would be very interesting and informative if these models can be cross-compared and a snapshot of their relative performance be provided. Such an objective can be achieved by setting up a common ground on which different models perform and compete, just like a tournament. Such a process is called benchmarking, two scenarios of which are set up here with one being microscopic and the other macroscopic.

16.3.1 Microscopic Benchmarking

The microscopic benchmarking employs a concrete example consisting of a set of hypothetical driving scenarios called regimes[1]. The purpose of microscopic benchmarking is to illustrate the performances 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 pre-determined and that of the follower is governed by a car-following model. Initially ($t = 0$), vehicle $i-1$ stands still at a downstream location 5000 meters referenced from an upstream location on the road ($x_{i-1}(0) = 5000$ m, $\dot{x}_{i-1}(0) = 0$ m/s, and $\ddot{x}_{i-1}(0) = 0$ m/s^2). Vehicle i, which is also still ($\dot{x}_i(0) = 0$ m/s and $\ddot{x}_i(0) = 0$ m/s^2), stands somewhere upstream which is to be determined and this location varies 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 $i-1$, i.e. $x_{i-1}(100) = 2810$ m, $\dot{x}_{i-1}(100) = 24$ m/s, and $\ddot{x}_{i-1}(100) = 0$ m/s^2. This change is designed to mimic the effect that a vehicle cuts off in front of another vehicle at 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 a deceleration process at a rate of $\ddot{x}_{i-1} = -3$ m/s^2 till 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/s^2 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 $\ddot{x}_{i-1} = -3$ m/s^2 till it comes to another full stop and remains there ever since. During all the time, the motion of the follower i is completely stipulated by a car-following model. The above scenario is formulated as follows.

[1]According to Merriam-Webster dictionary, a regime is the characteristic behavior or orderly procedure of a natural phenomenon or process

$$\begin{cases} x_{i-1} = 5000 \text{ m}, \dot{x}_{i-1} = 0 \text{ m/s}, \ddot{x}_{i-1} = 0 \text{ m/s}^2 & \text{when } 0 \leq t < 100 \text{ s} \\ x_{i-1} = 2810 \text{ m}, \dot{x}_{i-1} = 24 \text{ m/s} & \text{when } t = 100 \text{ s} \\ \ddot{x}_{i-1} = 0 \text{ m/s}^2 & \text{when } 100 \leq t < 200 \text{ s} \\ \ddot{x}_{i-1} = -3 \text{ m/s}^2 & \text{when } 200 \leq t < 208 \text{ s} \\ \ddot{x}_{i-1} = 0 \text{ m/s}^2 & \text{when } 208 \leq t < 300 \\ \ddot{x}_{i-1} = 2 \text{ m/s}^2 & \text{when } 300 \leq t < 318 \text{ s} \\ \ddot{x}_{i-1} = 0 \text{ m/s}^2 & \text{when } 318 \leq t < 400 \text{ s} \\ \ddot{x}_{i-1} = -3 \text{ m/s}^2 & \text{when } 400 \leq t < 412 \text{ s} \\ \ddot{x}_{i-1} = 0 \text{ m/s}^2 & \text{when } t \geq 412 \text{ s} \end{cases}$$

Driving regimes involved in the above example include:

- Start-up: vehicle i starts to move from stand-still when the scenario begins ($t > 0$ s).
- Speed-up: after start-up, vehicle i continues to accelerate to higher speeds ($0 < t < 100$ s).
- Free-flow: as vehicle i speeds up, it settles at its desired speed if it is unimpeded ($0 < t < 100$ s).
- Cut-off: a sudden decrease in spacing due to the new leader $i-1$ cutting off in front ($t = 100$ s).
- Following: vehicle i has to adopt $i-1$'s speed so as to follow the leader ($100 < t < 200$ s).
- Stop-and-go: vehicle i is forced to stop and go due to $i-1$'s brief stopping ($200 \geq t \leq 300$ s).
- Trailing: vehicle i is following a speeding leader ($300 < t < 400$ s).
- Approaching: vehicle i is getting close to a slower or stationary leader ($400 \geq t < 420$ s).
- Stopping: vehicle i tries to stop behind a stationary object separated by a minimum spacing ($t \geq 420$ s).

This example 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, reality check includes the following items:

- Start-up: whether the model itself is sufficient to start the vehicle up or an additional, external logic is needed.
- Speed-up: whether the model generates speed and acceleration profiles that make physical sense.
- Free-flow: whether the model settles at its desired speed without over-shooting or under-shooting.
- Cut-off: whether the model loses control or, if not, responds with reasonable control maneuver.
- Following: whether the model is able to adopt the leader's speed and follow the leader with 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 by its speeding leader, i.e., a vehicle is attracted to speed by its speeding leader.
- Approaching: whether the model is able to adjust the vehicle properly when inter-vehicle spacing closes up.
- Stopping: whether the model is able to stop the vehicle properly behind a stationary object, e.g. without over-shooting or under-shooting, separated by a minimum spacing, speed and acceleration dropping to zero 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 normally would back up a little under this situation and then identify a comfortable spacing to start car following.

16.3.2 Macroscopic Benchmarking

The macroscopic benchmarking employs a set of empirical data obtained from Georgia 400, a toll road with freeway by design located in Atlanta, GA, USA. The data contain one year worth of field observations at one

station across 4 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 16.5. This figure contains a set of 4 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-minute intervals. To highlight the average behavior of traffic flow, the cloud is further aggregated with respect to density and the result is shown as circles.

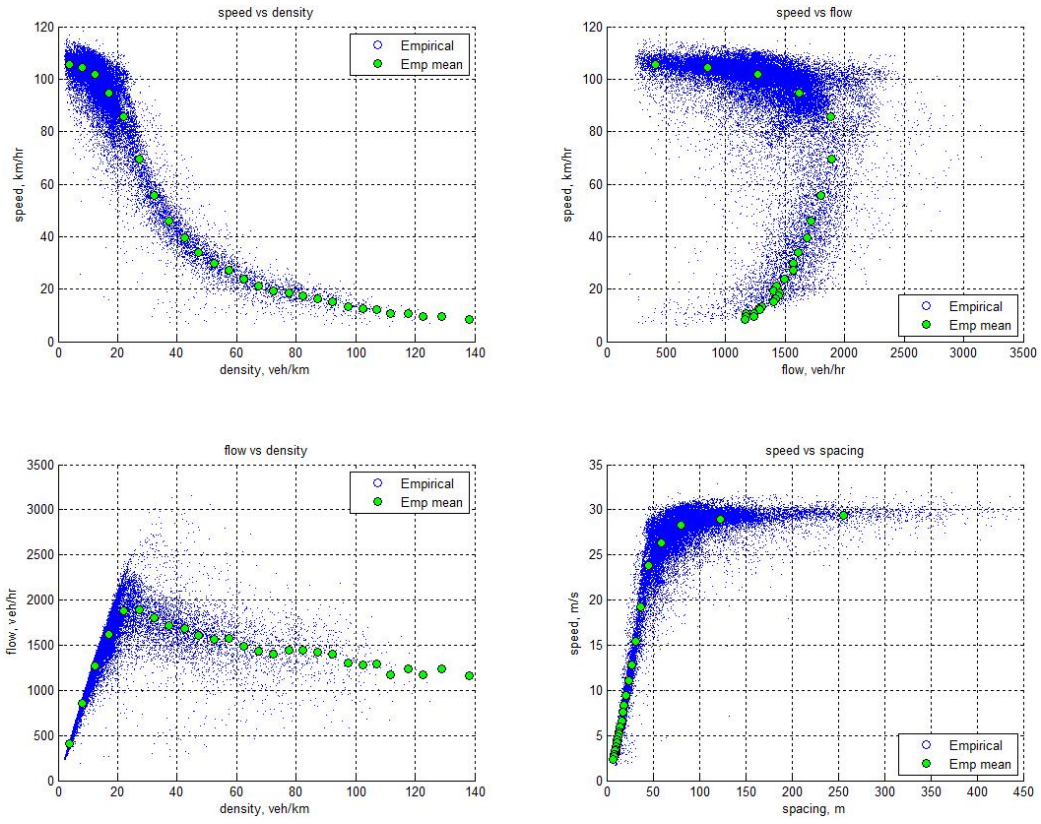


Figure 16.5: Empirical fundamental diagram observed from the field

Chapter 17

Pipes and Forbes Models

As the beginning of discussion on car-following models, this chapter introduces two simple models, i.e., Pipes model and Forbes model, both of which are derived from drivers' daily driving experiences.

17.1 Pipes Model

Pipes model [119] is based on a safe driving rule coined in California Motor 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 18.1 and putting the safety rule in mathematical language results in the following:

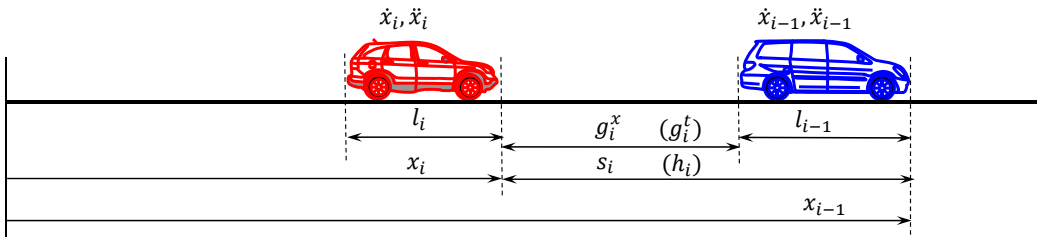


Figure 17.1: A car following scenario

$$g_i^x(t)_{min} = [(x_{i-1}(t) - x_i(t)) - l_{i-1}]_{min} = (s_i(t) - l_{i-1})_{min} = \frac{\dot{x}_i(t)}{0.447 \times 10} l_i \tag{17.1}$$

where $\dot{x}_i(t)$ is in m/s, and $g_i^x(t)$, $(x_{i-1}(t)$, and $x_i(t)$ is measured in meters. Assume a vehicle length of 6 meters, the model reduces to:

$$s_i(t)_{min} = 1.34\dot{x}_i(t) + 6 \tag{17.2}$$

or

$$h_i(t)_{min} = 1.34 + \frac{6}{\dot{x}_i(t)} \tag{17.3}$$

Equation 17.1, 17.2, or 17.3 constitutes the mathematical formulation of Pipes model.

17.1.1 Applications of Pipes Model

Pipes model can be applied in many ways, two foremost of which are automatic driving and computer simulation.

Automatic driving

Perhaps the simplest form of automatic driving is cruise control (CC). As an in-vehicle system, cruise control automatically controls the speed of a motor vehicle (by taking over the control of 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 vehicles keeps increasing on the road and the traffic becomes more and more crowded, the driver has to switch CC on and off so frequently that cruise control becomes less useful. In order to adapt to the dynamics of the vehicle in front, it is desirable that the cruise control system is able to adjust speed accordingly (rather than cruising at a constant speed) to maintain safe car following distance. Hence, adaptive or autonomous cruise control (ACC) system is developed. With the aid of distance sensors such as radar or laser, ACC allows the vehicle to slow down when approaching another vehicle and accelerate to the preset speed when traffic condition permits. To make this happen, the system requires an internal logic which relates vehicle speed to the distance between the vehicle and the one in front of it. Simple car-following models such as Pipes model

can be employed as the basis of such an internal logic. More specifically, Equation 17.1 can be rearranged as follows:

$$\dot{x}_i(t) \leq g_i^x(t)\frac{0.447 \times 10}{l_i} = 0.745 g_i^x(t) \tag{17.4}$$

the constant coefficient is resulted if a vehicle length of 6 meters applies. Therefore, the ACC 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.

Computer simulation

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, i.e. assign a value to each of them:

l_i length of vehicle i $\in$ {1,2,...,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 is 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

previous time step. The minimum safe spacing, $s_{min}(j,i)$ is determined according to California Motor Vehicle Code. Then, $s(j,i)$ is compared against $s_{min}(j,i)$. If $s(j,i)$ is less than $s_{min}(j,i)$, reduce the speed of the vehicle by ΔB_i, but don't go beyond 0. Otherwise, increase the speed of the vehicle by ΔA_i without exceeding its desired speed v_i. Then, 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 carry different objectives. The objective of automatic control is to guarantee safety yet achieving mobility (e.g., arriving at 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 realistic 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.

17.1.2 Properties of 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 or equivalently fundamental diagram.

Typically, the linkage between microscopic and macroscopic models can be addressed in two ways. One approach is to run simulation based on the microscopic model. Such a microscopic simulation typically involves random variables such as perception-reaction time, desired speed, acceleration rate, etc. 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, i.e., 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 steady state, any property of the system is unchanging in time. More specifically, a traffic system in steady state would consists of homogeneous vehicles which exhibit unform behavior over time and space. Therefore, under steady

state condition, vehicles lose their identities (e.g. $\tau_i \to \tau$ and $l_i \to l$), vehicles travel at uniform speed (i.e. $\dot{x_i} = \dot{x_j} \to v$ and $\ddot{x_i} \to 0$), drivers' desired speeds converge to free-flow speed (i.e. $v_i \to v_f$), and vehicle spacing $s_i(t)$ is replaced by the reciprocal of traffic density $\frac{1}{k}$. Hence, Pipes model reduces to:

$$\frac{1}{k} = 1.34v + 6 \text{ or } v = \frac{0.745}{k} - 4.47 \tag{17.5}$$

where k is measured in veh/m and v in m/s. The above speed-density relationship gives rise to the following flow-density and speed-flow relationships:

$$q = 0.745 - 4.47k \tag{17.6}$$

and

$$v = \frac{6q}{1 - 1.34q} \tag{17.7}$$

Equations 17.5, 17.6, and 17.7 constitute the mathematical representation of the fundamental diagram implied by Pipes model.

17.2 Forbes Model

Rather than ensuring safety distance between vehicles as Pipes model does, Forbes [38, 37] 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:

$$g_i^t(t) = h_i(t) - \frac{l_i}{\dot{x}_i} \geq \tau_i \tag{17.8}$$

Assume a reaction time of 1.5 seconds and vehicle length 6 m, the model becomes:

$$h_i(t) \geq 1.5 + \frac{6}{\dot{x}_i} \tag{17.9}$$

or

$$s_i(t) \geq 1.5\dot{x}_i + 6 \tag{17.10}$$

This is very similar to Pipes model except for a slight difference in the coefficient of the speed term which is interpreted as perception-reaction time τ_i. Therefore, Pipes model and Forbes model are essentially equivalent and can be generically expressed as:

$$s_i(t) \geq \tau_i\dot{x}_i + l_i \tag{17.11}$$

where τ_i and vehicle length l_i are model parameters. Note that applications and properties of Pipes model discussed above apply to Forbes model. In addition, the fundamental diagram implied by Pipes and Forbes models can be generically expressed as:

$$v = \frac{1}{\tau k} - \frac{l}{\tau} \tag{17.12}$$

$$q = \frac{1}{\tau} - \frac{l}{\tau}k \tag{17.13}$$

$$v = \frac{q1}{1 - \tau q} \tag{17.14}$$

where τ is average perception-reaction time and l is average vehicle length.

17.3 Benchmarking

Since Pipes and Forbes models are essentially equivalent, the following discussion addresses only Pipes model with the understanding that the result applies to Forbes model as well. Microscopic benchmarking refers to the scenario presented in 16.3.1 and macroscopic benchmarking refers to the scenario presented in 16.3.2.

17.3.1 Microscopic Benchmarking

For convenience, Pipes model is reproduced below:

$$\dot{x}_i(t + \Delta t) = \frac{s_i(t) - l_i}{\alpha} \tag{17.15}$$

where Δt is simulation time step and α is a constant resulted from unit conversion ($\alpha = 1.34$ if speed is in m/s and $l_i = 6$ m).

First, The model has problem with vehicle acceleration. Referring to the microscopic benchmarking scenario presented in 16.3.1, 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 Pipes model. A spacing of $s_i(0) = 5102$ m results in a speed of about 3803 m/s at the next time step (assuming $\Delta t = 1$ s), which requires an acceleration of 3803 m/s^2. It follows that an infinity speed and acceleration would be resulted if there is no leading vehicle in front. Therefore, the following external logic has to be imposed on Pipes model in order to limit its maximum acceleration:

$$\ddot{x}_i(t) = \frac{\dot{x}_i(t + \Delta t) - \dot{x}_i(t)}{\Delta t} \leq A_i \tag{17.16}$$

where A_i is the maximum acceleration of vehicle i, e.g. $A_i = 4$ m/s^2. With this addition, Pipes model loses its mathematical elegance which favors one-equation-for-all formulation. Even though an external logic is added, Pipes model still has problem with maximum speed. For example, it is true that the acceleration no longer exceeds A_i, but the vehicle can still reach unrealistic high speeds, e.g. $\dot{x}_i = 196$ m/s when $s_i = 590$ m. Therefore, another external logic has to be imposed to limit speed:

$$\dot{x}_i \leq v_i \tag{17.17}$$

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 $x_i = 8734$ m moving at speed $\dot{x}_i = 30$ m/s, while vehicle $i - 1$ stops at $x_{i-1} = 8762$ m. According to Pipes model, vehicle i's speed at the next step would be $\dot{x}_i \approx 16.42$ m/s. As such, the deceleration rate is $\ddot{x}_i = -13.58$ m/s^2. Hence, a third external logic has to be imposed to limit maximum deceleration B_i (e.g. -6 m/s^2):

$$\ddot{x}_i(t) = \frac{\dot{x}_i(t + \Delta t) - \dot{x}_i(t)}{\Delta t} \geq B_i \tag{17.18}$$

However, this addition introduces a new problem. For example, vehicle i's speed at the next step becomes $\dot{x}_i = 30 - 6 = 24$ m/s^2 and its location is $x_i = 8758$ m. This would leave a spacing of $s_i = 4$ m which is less than

a vehicle length $l_{i-1} = 6$ m, i.e., vehicle i has collided into vehicle $i-1$. Unfortunately, there is no easy remedy to the problem except for accepting the unrealistic deceleration behavior.

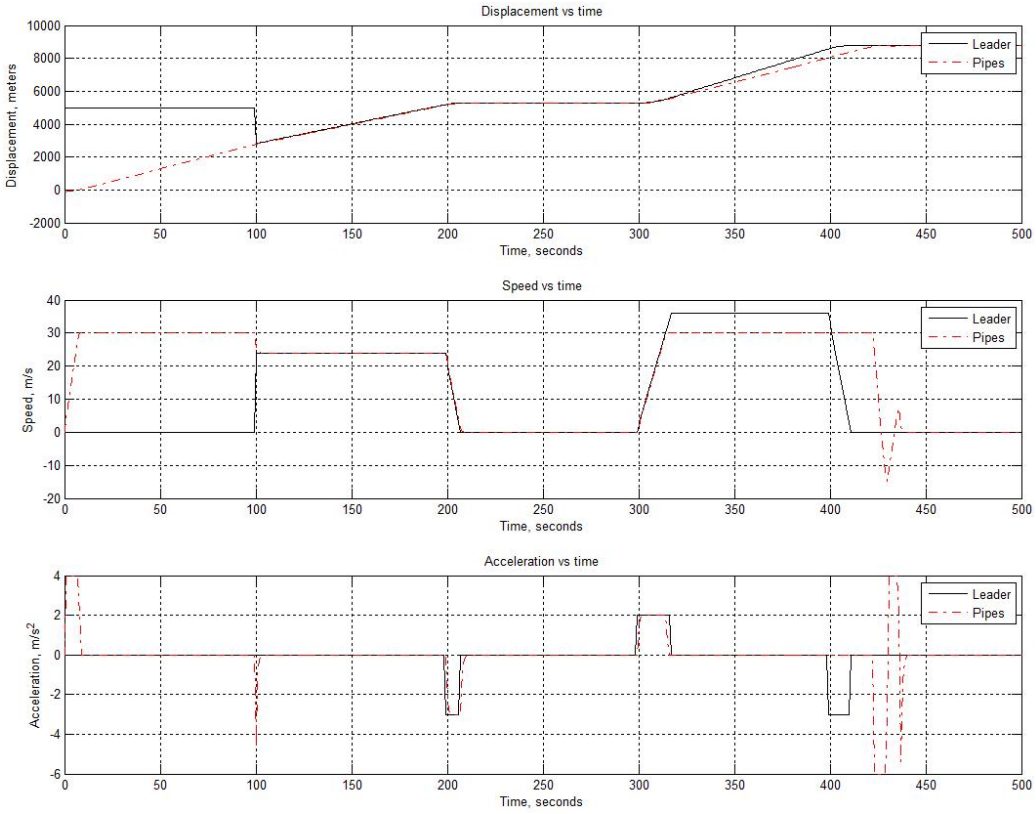


Figure 17.2: Microscopic benchmarking of Pipes model

Benchmarking result of Pipes model with constraints 17.16, 17.17, and 17.18 is plotted in Figure 17.2. 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 stand-still. See the figure when $t > 0$ s.
- Speed-up: 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 only available when a vehicle starts up. As the vehicle speeds up, acceleration decreases and eventually varnishes when the vehicle achieves its desired / cruising speed. See the figure when $0 < t < 100$ s.

- Free-flow: an external logic has to be imposed to limit maximum under free-flow condition. See the figure when $0 < t < 100$ s
- Cut-off: the model retains control and responds reasonably when a vehicle cuts in front. See the figure around $t = 100$ s.
- Following: the model is able to adopt the leader's speed and follow the leader by a reasonable distance. See the figure when $100 < t < 200$ s.
- Stop-and-go: the model is able to stop the vehicle safely behind its leader and start moving when the leader departs. See the figure when $200 \geq t \leq 300$ s.
- Trailing: the model is able to speed up normally without being tempted by its speeding leader. See the figure when $300 < t < 400$ s.
- Approaching: the model is unable to decelerate properly when approaching a stationary vehicle at a distance. The vehicle might collide into its leader when maximum deceleration is imposed. See the figure when $400 \geq t < 420$ s.
- Stopping: this portion is invalid since approaching fails. See the figure when $t \geq 420$ s.

The above benchmarking is based on the set of parameters in Table 17.1 and the outcome may vary under different set of parameters.

Table 17.1: Microscopic benchmarking parameters of 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

17.3.2 Macroscopic Benchmarking

The fundamental diagram implied by Pipes model is plotted in Figure 17.3 against empirical observations. The "cloud" contains 5-minute observations of flow, speed, and density, the circles are empirical observations aggregated with respect to density, and the curves are the equilibrium relationships implied by Pipes model.

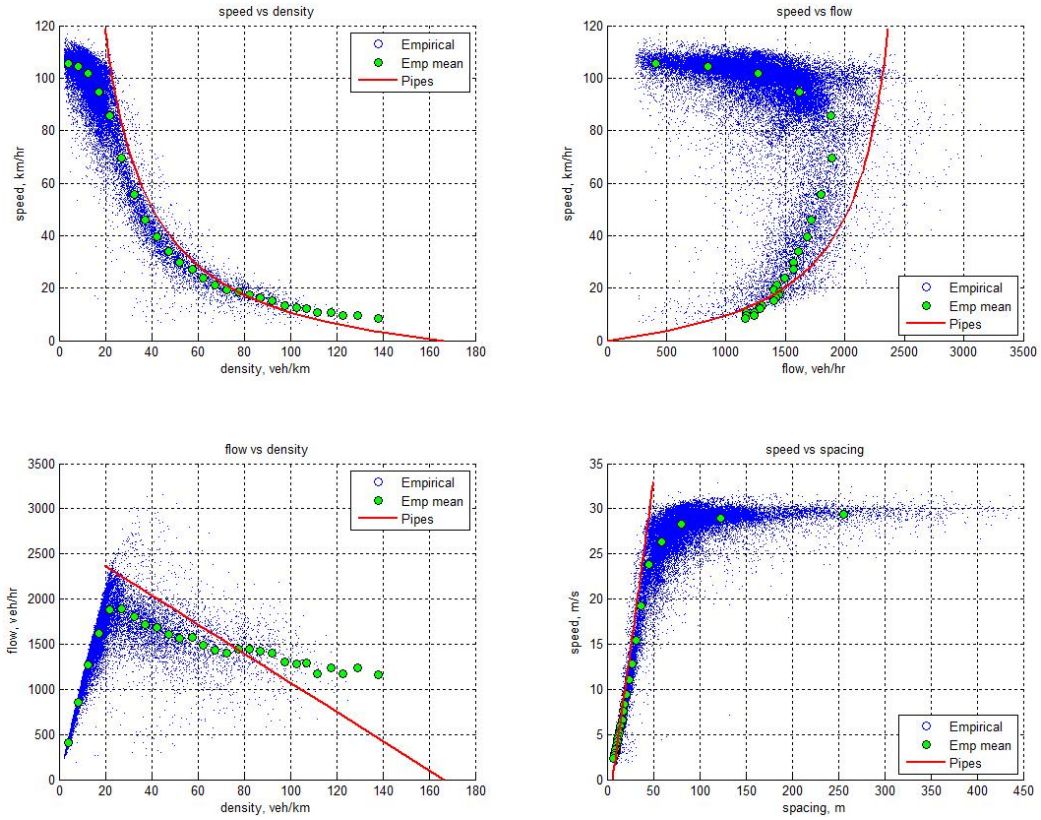


Figure 17.3: Fundamental diagram implied by Pipes model

These curves roughly fit the empirical data in the middle to upper range of density (e.g., $k > 20$ veh/km), but do not apply to the low density range (e.g., $k < 20$ veh/km). It appears that Pipes model is designed to literally describe car-following behavior. Cases when the leading vehicle is absent has to be handled by an external logic. In addition, 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 17.2 and the outcome may vary under different set of parameters.

Table 17.2: Macroscopic benchmarking parameters of Pipes model

α (τ)	l
1.34 s	6 m

Since Forbes model is essentially the same as Pipes model, the above benchmarking results apply to Forbes model as well.

Chapter 18

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's important is the role that General Motors played in the history of traffic flow theory. Back in 1950's, 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 for General Motors models (referred to as the GM models thereafter).

18.1 Development of GM Models

GM models [11, 43] 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:

$$response = f\ (sensitivity,\ stimuli)$$

When formulating the above relationship, GM researchers chose the subject vehicle's acceleration (deceleration is negative acceleration) issued after a reaction time, $\ddot{x}_i(t+\tau)$, as the response, see Figure 18.1. The consideration of stimuli and sensitivity evolved over time and resulted in a family of models.

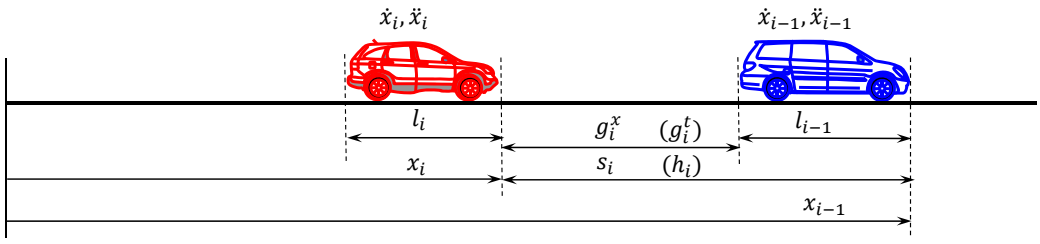


Figure 18.1: A car following scenario

18.1.1 GM1

Originally, GM 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:

$$\ddot{x}_i(t+\tau_i) = \alpha[\dot{x}_{i-1}(t) - \dot{x}_i(t)] \tag{18.1}$$

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/hr and $\dot{x}_{i-1}(t) = 100$ km/hr), 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/hr and $\dot{x}_{i-1}(t) = 120$ km/hr), the subject vehicle will accelerate because their 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/hr, $\dot{x}_{i-1}(t) = 100$ km/hr, and $s_i(t) = 50$ m
- Scenario 2: $\dot{x}_i(t) = 120$ km/hr, $\dot{x}_{i-1}(t) = 100$ km/hr, and $s_i(t) = 5000$ m

18.1.2 GM2

The effect of spacing motivated GM researchers to choose different sensitivity coefficients and, hence, the second generation model was resulted:

$$\ddot{x}_i(t+\tau_i) = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} [\dot{x}_{i-1}(t) - \dot{x}_i(t)] \tag{18.2}$$

Field experiments revealed that sensitivity coefficient α ranges between $(0.17, 0.74)$. 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.

18.1.3 GM3

The effect of spacing was partially address in GM2 because one has to frequently calibrate the sensitivity coefficient depending on car-following distances. The inconvenience seemed to suggest that spacing be explicitly included in the model, which led to the formulation of the third generation model:

$$\ddot{x}_i(t+\tau_i) = \alpha \frac{[\dot{x}_{i-1}(t) - \dot{x}_i(t)]}{[x_{i-1}(t) - x_i(t)]} \tag{18.3}$$

Having suppressed the issue of spacing, another problem pops up. The model is unable to predict any difference between the following scenarios:

- Scenario 1: In downtown Boston, one vehicle is following another at a spacing of 100 m with speeds $\dot{x}_i(t) = 30$ km/hr, $\dot{x}_{i-1}(t) = 10$ km/hr
- Scenario 2: On I-90, one vehicle is following another at a spacing of 100 m with speeds $\dot{x}_i(t) = 130$ km/hr, $\dot{x}_{i-1}(t) = 110$ km/hr

The subject driver on I-90 is certainly under a great deal of pressure to maintain safety during car following because, at such a high speed, a monent's lapse of attention would result in a catastrophe. In contrast, the subject driver in downtown Boston should have peace in 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/hr and the spacing is 100 m.

18.1.4 GM4

GM3's inability to differentiate high- and low-speed car following scenarios motivated GM 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 out of the sensitivity coefficient, leaving the remainder as new, dimensionless coefficient. This gives rise to the fourth generation model:

$$\ddot{x}_i(t+\tau_i) = \alpha \frac{\dot{x}_i(t+\tau_i)[\dot{x}_{i-1}(t) - \dot{x}_i(t)]}{[x_{i-1}(t) - x_i(t)]} \tag{18.4}$$

18.1.5 GM5

To generalize the results of the above GM models and, as it 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:

$$\ddot{x}_i(t+\tau_i) = \alpha \frac{[\dot{x}_i(t+\tau_i)]^m}{[x_{i-1}(t) - x_i(t)]^l}[\dot{x}_{i-1}(t) - \dot{x}_i(t)] \tag{18.5}$$

where x_i, $\dot{x}_i$, and $\ddot{x}_i$ are the displacement, speed, and acceleration of the subject vehicle i and similar notations apply 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.

18.2 Microscopic Benchmarking

The following segment of code implements GM4 model, a full-bloom one in the family. At time step j, the displacement, speed, and acceleration of each vehicle is 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, a: displacement, speed, and acceleration, respectively
i: vehicle ID $i \in$ {1,2,...,I}
j: time step $j \in$ {1,2,...J}
τ_i: perception-reaction time of driver i
Δt: simulation time step

Microscopic benchmarking refers to the scenario presented in 16.3.1 and benchmarking result of GM4 model is plotted in Figure 18.2 which is further elaborated as follows.

- Start-up: the model is unable to start a vehicle from stand-still. 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 desired speed v_i. Otherwise, vehicle i won't be able to reach that speed by itself. See the figure when $t > 0$ s.
- Speed-up: 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 the figure when $0 < t < 100$ s.
- Free-flow: the model predicts that vehicle i is unable to attain free-flow condition by itself unless it is set so by an external logic. As long as it follows slower leader, the model constantly decelerates the vehicle till it adopts the leader's speed. See the figure when $0 < t < 100$ s

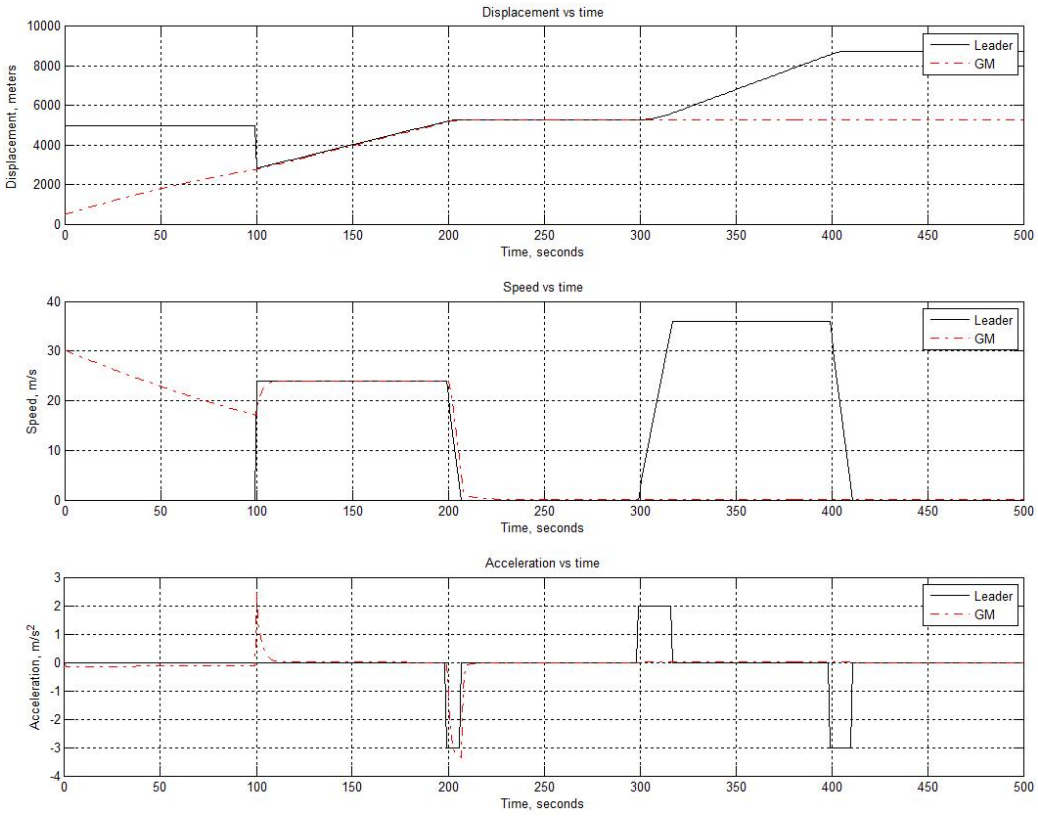


Figure 18.2: Microscopic benchmarking of GM4 model

- Cut-off: 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. See the figure around $t = 100$ s.
- Following: the model is able to adopt the leader's speed and follow the leader by a reasonable distance. See the figure when $100 < t < 200$ s.
- Stop-and-go: the model predicts that vehicle i will gradually but surely collide into its leader while maintaining a speed, regardless of how small the speed is. When the leader resumes motion, vehicle i will be stuck there because of its infinitesimally small speed. See the figure when $200 \geq t \leq 300$ s.
- Trailing: vehicle i is stuck and fails to catch up with its speeding leader, unless another external logic brings it out. See the figure when $300 < t < 400$ s. However, once the vehicle resumes motion, it will be tempted by its speeding leader and adopt the leader's speed. Such an effect is not shown in the figure.
- 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 18.1 and the outcome may vary under different set of parameters.

Table 18.1: Microscopic benchmarking parameters of GM4 model

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

18.3 Limitations of GM models

As a seminal work in the early history, GM models have spawned and inspired numerous research efforts that shaped today's traffic flow theory and, thus, the importance of this work can not be underestimated. Meanwhile, GM

models suffer some limitations which are presented below using GM4 as an example.

Universal car following On the one hand, GM4 model 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 would predict that a vehicle in Atlanta must be in car following with another vehicle in Boston even though they are over 1000 kilometers apart.

Attraction as a mechanism of motion If one compares GM4 (Equation 18.4) against Newton's law of universal gravitation (Equation 18.6) and Coulomb's law (Equation 18.7), one finds that they are strikingly similar to each other.

$$F = G\frac{m_1 m_2}{r^2} \tag{18.6}$$

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.

$$F = k_e\frac{q_1 q_2}{r^2} \tag{18.7}$$

where F is electrostatics 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 an equivalent of Coulomb's law as follows. Vehicle i will be repelled by its leader $i-1$ when i is approaching $i-1$ at a higher speed, while i will be attracted to $i-1$ should i falls 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, e.g., the very first vehicle to enter the highway? Then the subject vehicle could not start because there would be no one to pull it forward. Even if the subject vehicle is following a leader and the gap between them is opening, one doesn't feel like he or she is attracted to the leader. Actually, one speeds up because one would like to achieve his or her desired speed.

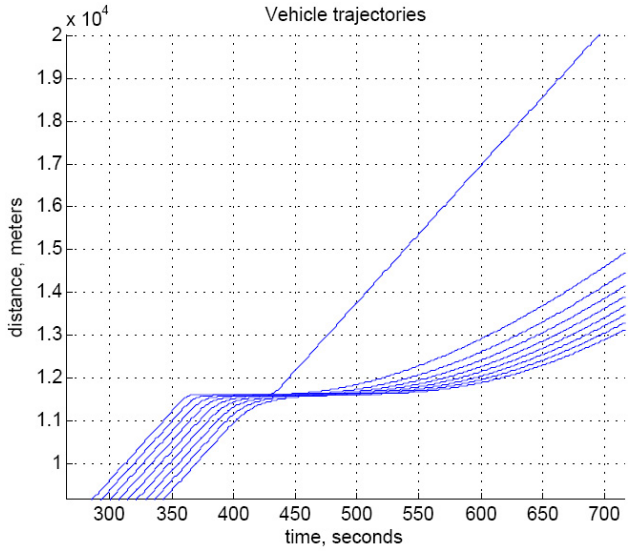


Figure 18.3: GM4 slow start

Slow start According to GM4, a vehicle at a stopped position is unable to get started. This is because the vehicle's speed at current step ($\dot{x}_i(t) = 0$) determines its acceleration in the next step ($\ddot{x}_i(t + \tau_i) = 0$), see Equation 18.4. Therefore, the vehicle has to maintain a non-zero 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 completely blocked by another vehicle on a highway. Otherwise, the subject vehicle has to slow down to an infinitesimal speed rather than 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 18.3 illustrates such a scenario where a noticeable gap is resulted between the first and second vehicle.

An intimate pair According to GM4, two vehicles can get arbitrarily close as long as they are traveling at the same speed, which is certainly not true - no one would dare to follow another with one-inch apart at 70 mph! 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.

Chapter 19

Micro-Macro Bridge

As mentioned in Chapter 17, 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 chapter is specifically devoted to such a purpose. It appears that GM5 model is ideal to serve as an 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:

$$\ddot{x}_i(t+\tau_i) = \alpha \frac{[\dot{x}_i(t+\tau_i)]^m}{[x_{i-1}(t) - x_i(t)]^l}[\dot{x}_{i-1}(t) - \dot{x}_i(t)] \tag{19.1}$$

In addition, those early equilibrium models are listed in Table 19.1:

Table 19.1: Single regime models

Author	Model	Parameters
Greenshields [50]	$v = v_f(1 - \frac{k}{k_j})$	v_f, k_j
Greenberg [49]	$v = v_m \ln(\frac{k_j}{k})$	v_m, k_j
Underwood [146]	$v = v_f e^{-\frac{k}{k_m}}$	v_f, k_m
Northwestern [29]	$v = v_f e^{-\frac{1}{2}(\frac{k}{k_m})^2}$	v_f, k_m
Drew [31]	$v = v_f[1 - (\frac{k}{k_j})^{n+\frac{1}{2}}]$	v_f, k_j, n
Pipes-Munjal [120, 97]	$v = v_f[1 - (\frac{k}{k_j})^n]$	v_f, k_j, n

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

19.1 Development of the Bridge

19.1.1 GM3 and Greenberg

If one chooses $m = 0$ and $l = 1$, GM5 reduces to GM3:

$$\ddot{x}_i(t + \tau_i) = \alpha \frac{[\dot{x}_{i-1}(t) - \dot{x}_i(t)]}{[x_{i-1}(t) - x_i(t)]} \tag{19.2}$$

Integrating both sides with regard to time t yields:

$$\dot{x}_i(t + \tau_i) = \alpha \ln |x_{i-1}(t) - x_i(t)| + C_i = \alpha \ln |s_i(t)| + C_i \tag{19.3}$$

where C_i is a constant of integration. Under steady-state or equilibrium condition, all vehicles are equally separated in space and traveling at the same speed. As such, driver identity i can be dropped because there is no difference among drivers. In addition, time t can be dropped because under equilibrium condition vehicle position and speed do not depend on time. Hence, $\dot{x}_i(t + \tau_i)$ is the same as traffic speed v. Moreover, spacing s becomes the reciprocal of traffic density k, i.e., $s \to \frac{1}{k}$. As traffic is jammed, $s \to l$, so $k \to k_j$. Therefore, Equation 19.3 can be reduced to:

$$v = \alpha \ln \frac{1}{k} + C \tag{19.4}$$

Field observations show that drivers tend to adjust their speeds according to prevailing traffic condition, e.g., density. Therefore, it is reasonable to assume a functional relationship between vehicle speed (also traffic speed under equilibrium condition) and traffic density, i.e. $v = V(k)$. It is recognized that traffic density varies between zero and jam density ($0 \leq k \leq k_j]$) and traffic speed also varies between zero and free-flow speed ($0 \leq v \leq v_f]$). It is also 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 level called critical density k_c. As traffic density continues to increase, traffic speed begins to drop until it reaches zero when density becomes maximum, i.e., k_j. The relationship between traffic speed and density can be roughly represented as the red solid curve in Figure

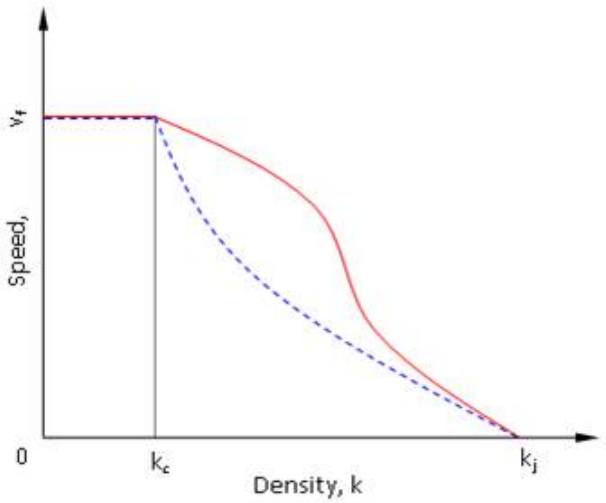


Figure 19.1: Speed-density relationship

19.1. Two known points on the speed-density curve are: (v_f, k_c) and $(0, k_j)$. Plug the two points into Equation 19.4, one obtains:

$$0 = \alpha \ln \frac{1}{k_j} + C \qquad\qquad C = \alpha \ln k_j$$

$$v_f = \alpha \ln \frac{1}{k_c} + \alpha \ln k_j = \alpha \ln \frac{k_j}{k_c}$$

$$\alpha = \frac{v_f}{\ln(k_j/k_c)}$$

Therefore

$$v = \frac{v_f}{\ln(k_j/k_c)} \ln \frac{k_j}{k} = v_m \ln \frac{k_j}{k}$$

if one defines $v_m = \frac{v_f}{\ln(k_j/k_c)}$. In summary, the speed-density relationship can be represented more explicitly as:

$$\begin{cases} v = v_f & \text{when } 0 \le k < k_c \\ v = v_m \ln \frac{k_j}{k} & \text{when } k_c \le k \le k_j \end{cases} \tag{19.5}$$

Note that the speed-density curve is fine tuned as the blue dashed curve in Figure 19.1. The astute reader may have recognized that the second branch of the above speed-density relationship is exactly Greenberg model.

Meanwhile, the corresponding flow-density relation can be derived from the above equation by using speed-flow-density relationship $q = kv$:

$$\begin{cases} q = v_f k & \text{when } 0 \leq k < k_c \\ q = v_m \ln\frac{k_j}{k} k & \text{when } k_c \leq k \leq k_j \end{cases} \tag{19.6}$$

In the first branch, flow q is a linear function of k and this holds true from $k = 0$ to $k = k_c$. In the second branch, the shape of the curve is not so clear. Observations show that as traffic density grows beyond k_c, flow continues to increase until it reaches the maximum flow called capacity, q_m, after which flow keeps decreasing until density reaches maximum, k_j, and essentially all vehicles are stopped, i.e., flow becomes zero. To find the capacity, one takes the first derivative of $q(k)$ and set it to zero:

$$\frac{dq}{dk} = \frac{v_f}{\ln(k_j/k_c)} \ln\frac{k_j}{k} + \frac{v_f}{\ln(k_j/k_c)} k \frac{k_j}{k} (-\frac{k_j}{k^2}) = \frac{v_f}{\ln(k_j/k_c)} (\ln\frac{k_j}{k} - 1) = 0$$

This gives,

$$\ln\frac{k_j}{k} - 1 = 0 \rightarrow k = k_m = \frac{k_j}{e}$$

where k_m is the density at capacity. Therefore, capacity q_m can be determined as:

$$q_m = \frac{v_f}{\ln(k_j/k_c)} \ln\frac{k_j}{k_m} k_m = \frac{v_f}{\ln(k_j/k_c)} \ln\frac{k_j}{\frac{k_j}{e}} \frac{k_j}{e} = \frac{v_f}{\ln(k_j/k_c)} \frac{k_j}{e}$$

To verify if q_m does exist, take the second derivative of q with respect to k:

$$\frac{d^2q}{dk^2} = \frac{v_f}{\ln(k_j/k_c)} \ln\frac{k}{k_j} = -\frac{v_f}{\ln(k_j/k_c)} \frac{1}{k}$$

where $k_j > k_m = k_j/e > k_c$, so $\ln(k_j/k_c) \geq \ln(e) > 0$ and by definition $k > 0$, so the second derivative of q is negative, which means q_m does exist. Therefore, the flow-density relationship can be plotted as the green solid

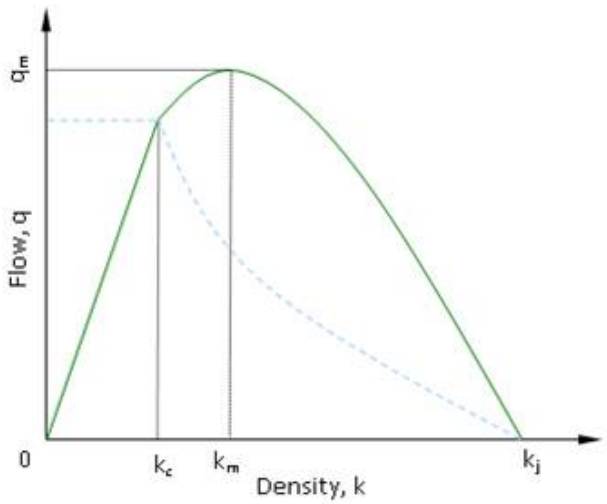


Figure 19.2: Flow-density relationship

curve in Figure 19.2. For easy reference, the speed-density curve is plotted in the background to show how it relates to the flow-density curve.

The above relation between Greenberg and GM3 suggests that it might be possible 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.

19.1.2 Greenshields Model

Setting $l = 2$ and $m = 0$ in GM5 (Equation 19.1) yields:

$$\ddot{x}_i(t + \tau_i) = \alpha \frac{[\dot{x}_{i-1}(t) - \dot{x}_i(t)]}{[x_{i-1}(t) - x_i(t)]^2} \tag{19.7}$$

where all variables are defined before. Integrate both sides with respect to time t:

$$\dot{x}_i(t + \tau_i) = -\alpha \frac{1}{[x_{i-1}(t) - x_i(t)]} + C \tag{19.8}$$

where C is a constant of integration. Under equilibrium condition, the above equation becomes:

$$v = -\alpha k + C \tag{19.9}$$

Consider boundary condition $k = 0 \Rightarrow v = v_f$. Plug in the above equation: $v_f = -\alpha 0 + C$. Therefore, $C = v_f$. In addition, when $k = k_j \Rightarrow v = 0$, so $0 = -\alpha k_j + v_f$. Hence, $\alpha = v_f/k_j$. Putting everything together, Greenshields model is resulted:

$$v = v_f - \frac{v_f}{k_j}k \tag{19.10}$$

19.1.3 Underwood Model

Setting $l = 2$ and $m = 1$ in GM5 and following similar procedure to the above:

$$\ddot{x}_i(t+\tau_i) = \alpha \frac{\dot{x}_i(t+\tau_i)[\dot{x}_{i-1}(t) - \dot{x}_i(t)]}{[x_{i-1}(t) - x_i(t)]^2} \tag{19.11}$$

$$\frac{\ddot{x}_i(t+\tau_i)}{\dot{x}_i(t+\tau_i)} = \alpha \frac{[\dot{x}_{i-1}(t) - \dot{x}_i(t)]}{[x_{i-1}(t) - x_i(t)]^2} \tag{19.12}$$

$$\ln(\dot{x}_i(t+\tau_i)) = -\alpha \frac{1}{[x_{i-1}(t) - x_i(t)]} \tag{19.13}$$

Under equilibrium condition, the above equation becomes:

$$\ln(v) = -\alpha k + C \tag{19.14}$$

$$v = e^{-\alpha k + C} = C_1 e^{-\alpha k} \tag{19.15}$$

When $k = 0$, $v = v_f$. Therefore, $v = v_f = C_1 e^{-\alpha \times 0} = C_1$. Hence,

$$v = v_f e^{-\alpha k} \tag{19.16}$$

Note that, though this equation takes the form of Underwood model, coefficient α has yet to be determined and the equation loses meaning when $k = k_j$. A similar treatment to Greenberge's model may be considered. Let $v_m = v_f/e$ and denote $k_m = k(v_m)$. One obtains

$$v_m = \frac{v_f}{e} = \frac{v_f}{e^{\alpha k_m}} \tag{19.17}$$

Hence, $\alpha k_m = 1$ or $\alpha = 1/k_m$. Therefore, Underwood model is resulted:

$$v = v_f e^{-k/k_m} \tag{19.18}$$

Actually, v_m and k_m are the optimal speed and density implied by Underwood model To verify, one formulates flow $q = kv$ and find its maximum value:

$$q = kv = kv_f e^{-k/k_m} \tag{19.19}$$

Take derivative of q with respective to k and set the result to zero:

$$\frac{dq}{dk} = v + k\frac{dv}{dk} = v_f e^{-k/k_m} + kv_f \frac{-1}{k_m} e^{-k/k_m} = 0 \tag{19.20}$$

Solving the above equation yields optimal values $k^* = k_m$ and $v^* = v_m = v_f/e$ and $q^* = q_m = k_m v_f/e$.

19.1.4 Drake (Northwestern) Model

Setting $l = 3$ and $m = 1$ in GM5 and following similar procedure to the above:

$$\ddot{x}_i(t+\tau_i) = \alpha \frac{\dot{x}_i(t+\tau_i)[\dot{x}_{i-1}(t) - \dot{x}_i(t)]}{[x_{i-1}(t) - x_i(t)]^3} \tag{19.21}$$

$$\frac{\ddot{x}_i(t+\tau_i)}{\dot{x}_i(t+\tau_i)} = \alpha \frac{[\dot{x}_{i-1}(t) - \dot{x}_i(t)]}{[x_{i-1}(t) - x_i(t)]^3} \tag{19.22}$$

$$\ln(\dot{x}_i(t+\tau_i)) = -\frac{1}{2}\alpha \frac{1}{[x_{i-1}(t) - x_i(t)]^2} + C \tag{19.23}$$

$$\ln(v) = -\frac{1}{2}\alpha k^2 + C \text{ or } v = C_1 e^{-\alpha k^2/2} \tag{19.24}$$

When $k = 0$, $v = v_f$. One obtains $v = v_f = C_1 e^{-\alpha \times 0} = C_1$. Hence

$$v = v_f e^{-\alpha k^2/2} \tag{19.25}$$

Notice that, though this equation takes the form of Drake model, coefficient α has yet to be determined and the equation loses meaning when $k = k_j$. A similar treatment to Greenberge's model may be considered. Set

$v_m = v_f e^{-1/2}$ and denote $k_m = k(v_m)$. Plugging v_m and k_m into the above equation yields $\alpha = k_m$. Therefore, Drake model is resulted:

$$v = v_f e^{-\frac{1}{2}(\frac{k}{k_m})^2} \tag{19.26}$$

To verify that v_m and k_m are optimal values, take the first derivative of $q = kv$ and set the result to zero:

$$\frac{dq}{dk} = v + k\frac{dv}{dk} = v_f e^{-\frac{1}{2}(\frac{k}{k_m})^2} + k(-\frac{k}{k_m^2})v_f e^{-\frac{1}{2}(\frac{k}{k_m})^2} = 0 \tag{19.27}$$

This confirms that optimal k^* and v^* are indeed k_m and v_m, respectively, and the capacity implied by Drake model is $q_m = k_m v_m = k_m v_f e^{-1/2}$.

19.1.5 Pipes-Munjal model

Setting $l = n + 1$ and $m = 0$ in GM5 yields:

$$\ddot{x}_i(t + \tau_i) = \alpha \frac{[\dot{x}_{i-1}(t) - \dot{x}_i(t)]}{[x_{i-1}(t) - x_i(t)]^{n+1}} \tag{19.28}$$

Integrate both sides with respect to time t:

$$\dot{x}_i(t + \tau_i) = -\frac{\alpha}{n} \frac{1}{[x_{i-1}(t) - x_i(t)]^n} + C \tag{19.29}$$

where C is a constant of integration. Under equilibrium condition, the above equation becomes:

$$v = -\frac{\alpha}{n} k^n + C \tag{19.30}$$

Plugging boundary condition $k = 0 \Rightarrow v = v_f$ into the above equation yields: $v_f = -\frac{\alpha}{n}0 + C$. Therefore, $C = v_f$. In addition, $k = k_j \Rightarrow v = 0$, so $0 = -\frac{\alpha}{n}k_j^n + v_f$. Hence, $\alpha = nv_f/k_j^n$. Putting everything together, Pipes-Munjal model is resulted:

$$v = v_f[1 - (\frac{k}{k_j})^n] \tag{19.31}$$

19.1.6 Drew model

Since Drew model and 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 Drew model. Hence, Drew model corresponds to GM5 with $l = n + 1\frac{1}{2}$ and $m = 0$.

19.1.7 Summary of the bridge

Summarizing the above, a diagram can be drawn that relates models discussed above to GM5. Figure 19.3 serves such a purpose with horizontal axis as speed exponent m and vertical axis as spacing exponent l of GM5. Macroscopic equilibrium models are labeled in red and microscopic car-following models are in blue. Circles on grid denote models and their corresponding m and l combination in relation to GM5.

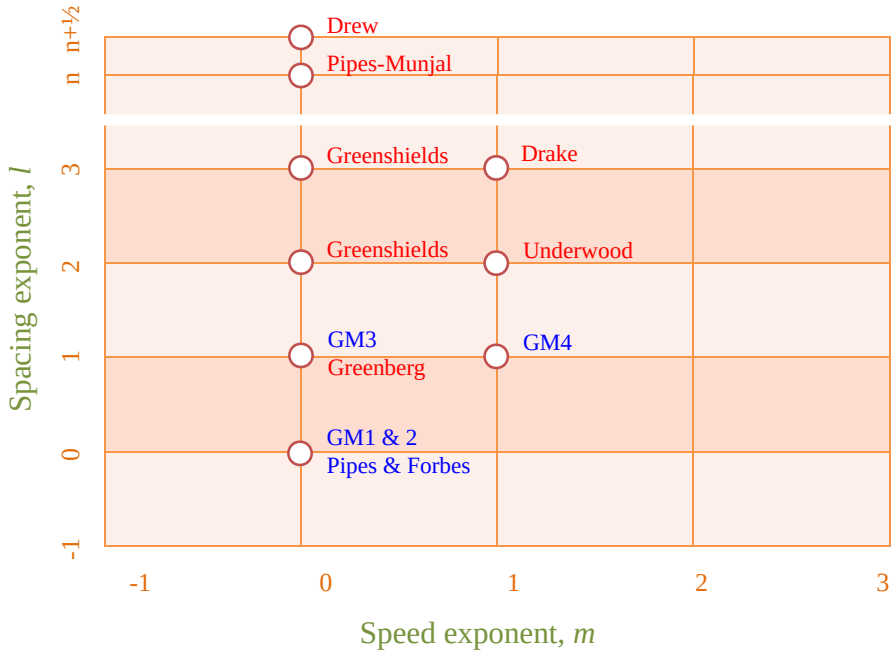


Figure 19.3: Micro-macro bridge

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 $\alpha = 1.36$ and $C = l_i$, one obtains Pipes model, while $\alpha = \tau_i$ and and $C = l_i$ leads to Pipes model.

19.2 Macroscopic Benchmarking

Macroscopic benchmarking refers to the scenario presented in 16.3.2. Fundamental diagrams implied by GM models and their associated equilibrium models are presented in Figure 19.4 against empirical observations. It can be seen that these fundamental diagrams achieve varying success in fitting empirical data, but none of them fits the data reasonable well in the entire range of density. For example, Greenshields model over-predicts speed (and hence flow) in the majority density range except for free-flow (i.e. low density) condition; Greenberg model has problem fitting the data under free-flow condition; Underwood model, perhaps the best among the list, under-estimates speed at low densities and over-estimate speed at high densities, and capacity occurs at much lower speed than it ought to be; Drake (Northwestern) model has an flow-density curve that is convex in high density range; Drew and Pipes-Munjal model, which are essentially the same but are shown slightly different to avoid complete overlap, share the same problem as that of Greenshields' but to a less extend.

The above benchmarking is based on the set of parameters in Table 19.2 and the outcome may vary under different set of parameters

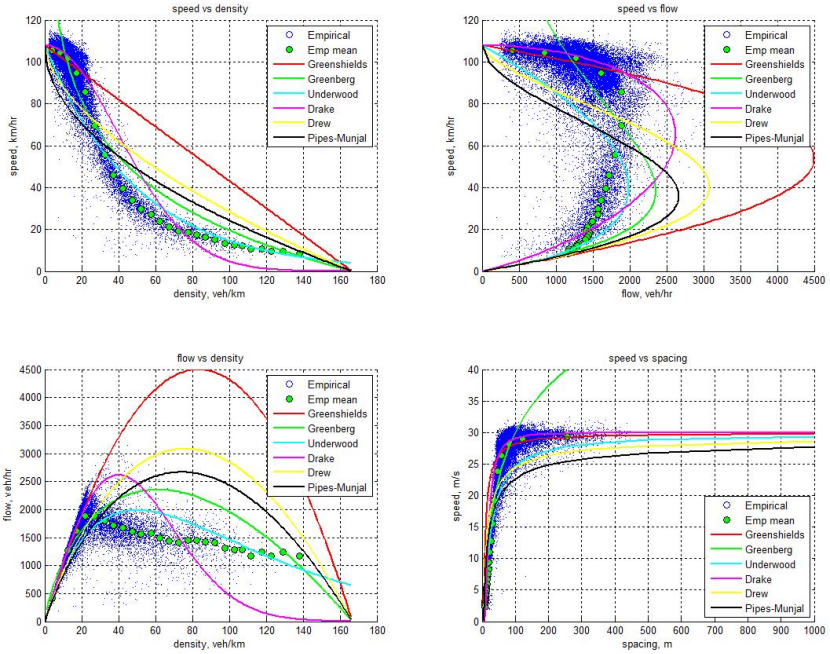


Figure 19.4: Fundamental diagrams implied GM and associated models

Table 19.2: Macroscopic benchmarking parameters of GM associated models

Greenshields	v_f	k_j	-
	30 m/s	1/6 veh/m	-
Greenberg	v_m	k_j	k_c
	10.7 m/s	1/6 veh/m	0.01 veh/m
Underwood	v_f	k_m	-
	30 m/s	0.05 veh/m	-
Drake (NW)	v_f	k_m	n
	30 m/s	0.04 veh/m	2
Drew	v_f	k_j	n
	30 m/s	1/6 veh/m	0.1
Pipes-Munjal	v_f	k_j	n
	30 m/s	1/6 veh/m	0.5

Chapter 20

Gipps Model

Pipes, Forbes, and GM models introduced in previous chapters are all single-regime models, i.e. they have only one equation that applies to the entire driving process without worrying about 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 a concern. For example, a driver may encounter different regimes such as start-up, speed-up, free-flow, cut-off, following, stop-and-go, trailing, approaching, and stopping. A one-equation model may or may not apply to all regimes. As such, multi-regime models might be helpful in capturing different driving scenarios. This chapter introduces a model along this line - the Gipps model.

20.1 Model Formulation

Just like Pipes and Forbes models, Gipps [46] model is derived from a safety driving rule, perhaps a more realistic but conservative one. A driver typically employ a safety rule to evaluate if the current car-following situation is safe. For example,

- Pipes rule stipulates that, at each moment, the driver need to estimate his or her own speed (in mph), divide it by 10 and multiply the quotient by a car length, and the result is the minimum gap to maintain. If the actual gap is shorter, fall back; otherwise, it is safe.
- Forbes rule ensures safety headway. For example, a driver with a perception-reaction time of 3 seconds can start counting "one thousand

one, one thousand two, one thousand three" when the leader passes a roadside utility pole. If the subject driver passes the pole before the counting is finished, a 3-second headway is not maintained; otherwise, it is safe.

Though the above two safety rules may sound reasonable to 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 should the vehicle ahead came to a sudden stop" [46]. Put it in another way, at any moment, the following driver should leave enough safe distance in front such that, in case the leading vehicle commences emergency brake, the subject driver has time to responde and decelerate to a stop behind the leading vehicle without a collision. Gipps car-following model is based on such an assumption and the scenario is depicted in Figure 28.1.

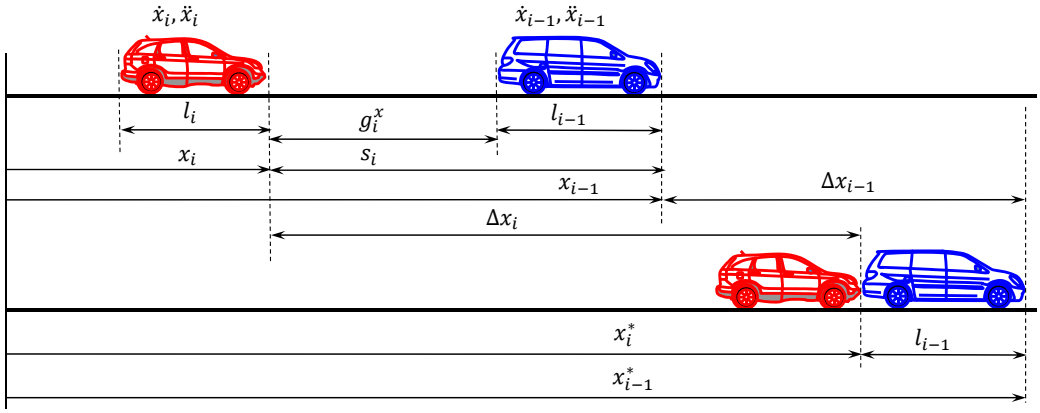


Figure 20.1: Gipps car following scenario

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 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 in a duration of τ_i, trying to understand the situation, evaluate potential options, and then decide to apply brake as well at a *tolerable* rate 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:

$$x_{i-1}^* = x_{i-1}(t) - \frac{\dot{x}_{i-1}^2(t)}{2B_{i-1}} \tag{20.1}$$

Meanwhile, vehicle i travels a distance of $\frac{\dot{x}_i(t)+\dot{x}_i(t+\tau_i)}{2}\tau_i$ during 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:

$$x_i^* = x_i(t) + \frac{\dot{x}_i(t)+\dot{x}_i(t+\tau_i)}{2}\tau_i - \frac{\dot{x}_i^2(t+\tau_i)}{2b_i} \tag{20.2}$$

To be conservative, Gipps added one more buffer space term in the above equation:

$$x_i^* = x_i(t) + \frac{\dot{x}_i(t)+\dot{x}_i(t+\tau_i)}{2}\tau_i + \dot{x}_i(t+\tau_i)\theta - \frac{\dot{x}_i^2(t+\tau_i)}{2b_i} \tag{20.3}$$

where θ is an extra buffer time appended to the perception-reaction time. To ensure safety, the following relationship must hold:

$$x_{i-1}^* - l_{i-1} \geq x_i^* \tag{20.4}$$

Plugging in everything, the above inequality translates to:

$$x_{i-1}(t) - \frac{\dot{x}_{i-1}^2(t)}{2B_{i-1}} - l_{i-1} \geq x_i(t) + \frac{\dot{x}_i(t)+\dot{x}_i(t+\tau_i)}{2}\tau_i + \dot{x}_i(t+\tau_i)\theta - \frac{\dot{x}_i^2(t+\tau_i)}{2b_i} \tag{20.5}$$

which corresponds to a spacing of $s_i(t) = x_{i-1}(t) - x_i(t)$, i.e.

$$s_i(t) \geq \frac{\dot{x}_i(t)+\dot{x}_i(t+\tau_i)}{2}\tau_i + \dot{x}_i(t+\tau_i)\theta - \frac{\dot{x}_i^2(t+\tau_i)}{2b_i} + \frac{\dot{x}_{i-1}^2(t)}{2B_{i-1}} + l_{i-1} \tag{20.6}$$

Though the above inequality can serve the need of 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 inputs and which ones are output. Our daily driving experiences show the following: $s_i(t)$ and l_{i-1} can be visually measurable though not humans are

so precise; $\dot{x}_i(t)$ and $\dot{x}_{i-1}(t)$ are measurable from speedometer or motion relative to 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 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:

$$-\frac{1}{2b_i}\dot{x}_i^2(t+\tau_i)+(\frac{\tau_i}{2}+\theta)\dot{x}_i(t+\tau_i)+\frac{\dot{x}_i(t)\tau_i}{2}+\frac{\dot{x}_{i-1}^2(t)}{2B_{i-1}}+l_{i-1}-s_i(t)\leq 0 \quad (20.7)$$

The roots of the above quadratic equation are:

$$\dot{x}_i(t+\tau_i)=b_i(\frac{\tau_i}{2}+\theta)\pm\sqrt{b_i^2(\frac{\tau_i}{2}+\theta)^2-b_i[-\dot{x}_i(t)\tau_i-\frac{\dot{x}_{i-1}^2(t)}{B_{i-1}}-2l_{i-1}+2s_i(t)]} \quad (20.8)$$

Let $\theta=\tau_i/2$ as suggested by Gipps:

$$\dot{x}_i(t+\tau_i)=-b_i\tau_i\pm\sqrt{b_i^2\tau_i^2-b_i[-\dot{x}_i(t)\tau_i-\frac{\dot{x}_{i-1}^2(t)}{B_{i-1}}-2l_{i-1}+2s_i(t)]} \quad (20.9)$$

Consider the signs of the roots and that speed is a positive value, the solution to inequality (20.7) is:

$$0\leq\dot{x}_i(t+\tau_i)\leq-b_i\tau_i+\sqrt{b_i^2\tau_i^2-b_i[-\dot{x}_i(t)\tau_i-\frac{\dot{x}_{i-1}^2(t)}{B_{i-1}}-2l_{i-1}+2s_i(t)]} \quad (20.10)$$

The above derivation has formulated Gipps model in car-following regime. In free-flow regime, i.e., vehicle i is not blocked by a leading vehicle, Gipps suggests the following speed choice:

$$0\leq\dot{x}_i(t+\tau_i)\leq\dot{x}_i(t)+2.5A_i\tau_i(1-\frac{\dot{x}_i(t)}{v_i})\sqrt{0.025+\frac{\dot{x}_i(t)}{v_i}} \quad (20.11)$$

where A_i is the maximum acceleration that driver i is willing to apply and v_i desirable speed that the driver i is willing to travel wherever possible.

Unlike inequality (20.10) which is derived from a safety rule, this speed choice model is an empirical one obtained from fitting vehicle experiment data. It basically accelerates/decelerates the vehicle toward the desirable speed without causing oscillation.

The above two speed choices may bring about a little confusion when they are applied to vehicle control because one has to decide which choice to make. For example, one should choose (20.10) in case of car following and (20.11) in case of free flow. However, where is the cut off point between free flow and car following? To save this confusion, Gipps suggests that there is no need to make the distinction. Under any situation, one just need to compute the two speed choices and take the lesser one, i.e.,

$$\dot{x}_i(t+\tau_i) = \min \begin{cases} \dot{x}_i(t) + 2.5A_i\tau_i(1-\frac{\dot{x}_i(t)}{v_i})\sqrt{0.025+\frac{\dot{x}_i(t)}{v_i}} & \text{FF} \\ -b_i\tau_i + \sqrt{b_i^2\tau_i^2 - b_i[\dot{x}_i(t)\tau_i - \frac{\dot{x}_{i-1}^2(t)}{B_{i-1}} + 2l_{i-1} - 2s_i(t)]} & \text{CF} \end{cases} \tag{20.12}$$

note FF: free-flow condition; CF: car-following condition.

20.2 Properties of Gipps model

As usual, the macroscopic property of Gipps model under equilibrium condition is of primary interest. To simplify analysis, Gipps safety rule presented in (20.6) can be re-written as follows if one ignores the speed change during perception-reaction process and the additional buffer time θ:

$$s_i(t) \geq \dot{x}_i(t)\tau_i - \frac{\dot{x}_i^2(t)}{2b_i} + \frac{\dot{x}_{i-1}^2(t)}{2B_{i-1}} + l_{i-1} \tag{20.13}$$

Setting both sides equal and rearranging terms yield

$$s_i(t) = \dot{x}_i(t)\tau_i - \frac{\dot{x}_i^2(t)}{2b_i} + \frac{\dot{x}_{i-1}^2(t)}{2B_{i-1}} + l_{i-1} \tag{20.14}$$

Under equilibrium conditions and take the equal sign, the above car-following model leads to the following density-speed relationship:

$$\frac{1}{k} = (-\frac{1}{2b} + \frac{1}{2B})v^2 + \tau v + l \tag{20.15}$$

or

$$\frac{1}{k} = \gamma v^2 + \tau v + l \tag{20.16}$$

where k is traffic density, $\gamma = -\frac{1}{2b} + \frac{1}{2B}$, $b < 0$ is average tolerable braking rate, $B < 0$ is average emergency braking rate, v average traffic speed, τ average perception-reaction time, and l average nominal vehicle length. The corresponding flow-speed relationship is:

$$q = \frac{v}{\gamma v^2 + \tau v + l} \tag{20.17}$$

To find capacity, one takes the first derivative of flow q with respect to v and set the result to zero:

$$\left.\frac{dq}{dv}\right|_{v_m} = -\left.\frac{\gamma - \frac{l}{v^2}}{(\gamma v + \tau' + \frac{l}{v})^2}\right|_{v_m} = 0 \tag{20.18}$$

Solving the equation yields

$$v_m = \sqrt{\frac{l}{\gamma}} \tag{20.19}$$

and correspondingly,

$$q_m = \frac{1}{2\sqrt{\gamma l} + \tau} \tag{20.20}$$

After checking the second derivative of q at v_m, it turns out that q_m is indeed the maximum value of v.

20.3 Benchmarking

Microscopic benchmarking refers to the scenario presented in 16.3.1 and macroscopic benchmarking refers to the scenario presented in 16.3.2.

20.3.1 Microscopic Benchmarking

Benchmarking result of Gipps model is plotted in Figure 20.2. The performance of Gipps model is summarized as follows:

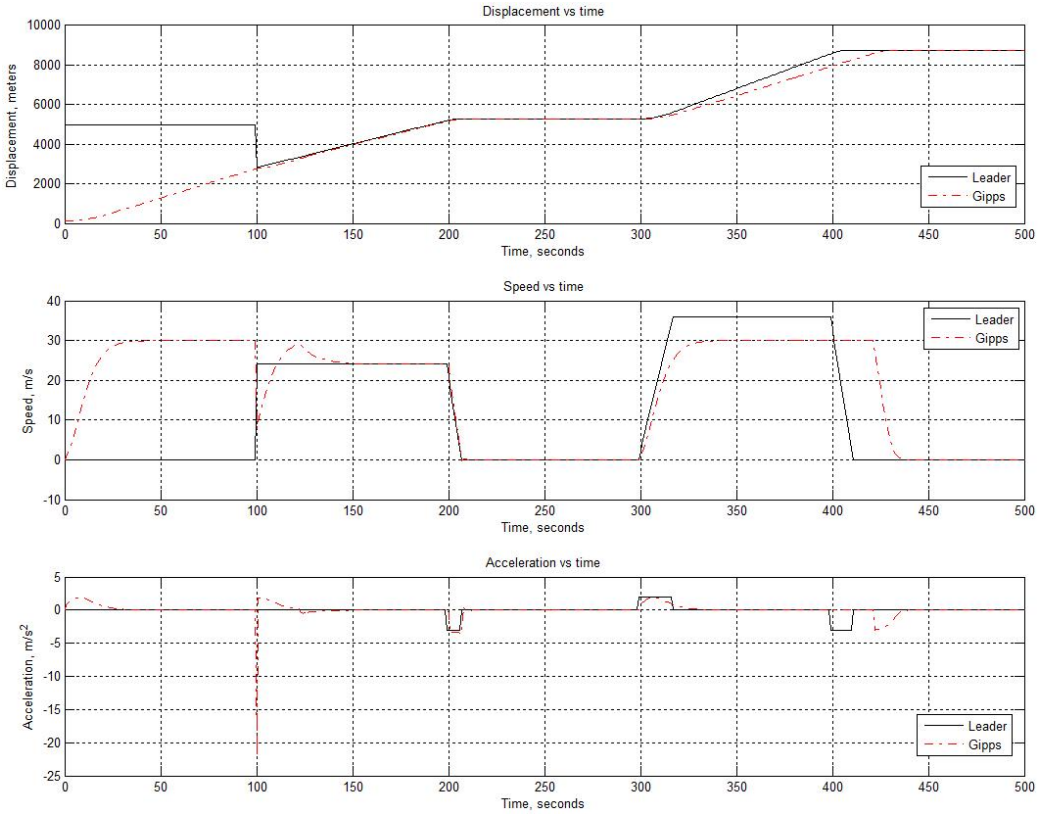


Figure 20.2: Microscopic benchmarking of Gipps model

- Start-up: the model is able to start the vehicle up from stand-still. See the figure when $t > 0$ s.
- Speed-up: the model is able to speed the vehicle up realistically to its desired speed. See the figure when $0 < t < 100$ s.
- Free-flow: the model is able to reach and settle at desired speed under free-flow condition. See the figure when $0 < t < 100$ s
- Cut-off: the model over-decelerates slightly which causes a small oscillation in speed, but in general the model retains control and responds reasonably when a vehicle cuts in front. See the figure around $t = 100$ s.
- Following: the model is able to adopt the leader's speed and follow the leader by a reasonable distance. See the figure when $100 < t < 200$ s.
- Stop-and-go: the model is able to stop the vehicle safely behind its leader and start moving when the leader departs. See the figure when $200 \geq t \leq 300$ s.
- Trailing: the model is able to speed up normally without being tempted by its speeding leader. See the figure when $300 < t < 400$ s.
- Approaching: the model is able to decelerate properly when approaching a stationary vehicle at a distance. See the figure when $400 \geq t < 420$ s.
- Stopping: the model is able to stop the vehicle safely behind the stationary vehicle. See the figure when $t \geq 420$ s.

The above benchmarking is based on the set of parameters in Table 20.1 and the outcome may vary under different set of parameters.

Table 20.1: Microscopic benchmarking parameters of Gipps model

l_i	v_i	τ_i	α	b_i
6 m	30 m/s	1.0 s	1.34	-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

20.3.2 Macroscopic Benchmarking

The fundamental diagram implied by Gipps model is plotted in Figure 20.3 against empirical observations. Model parameters are set the same as suggested in the original paper. It can be seen that the model fits empirical data reasonably well except for free-flow conditions (i.e. in the low density range). In addition, the speed at capacity predicted by Gipps model is much lower than it should be.

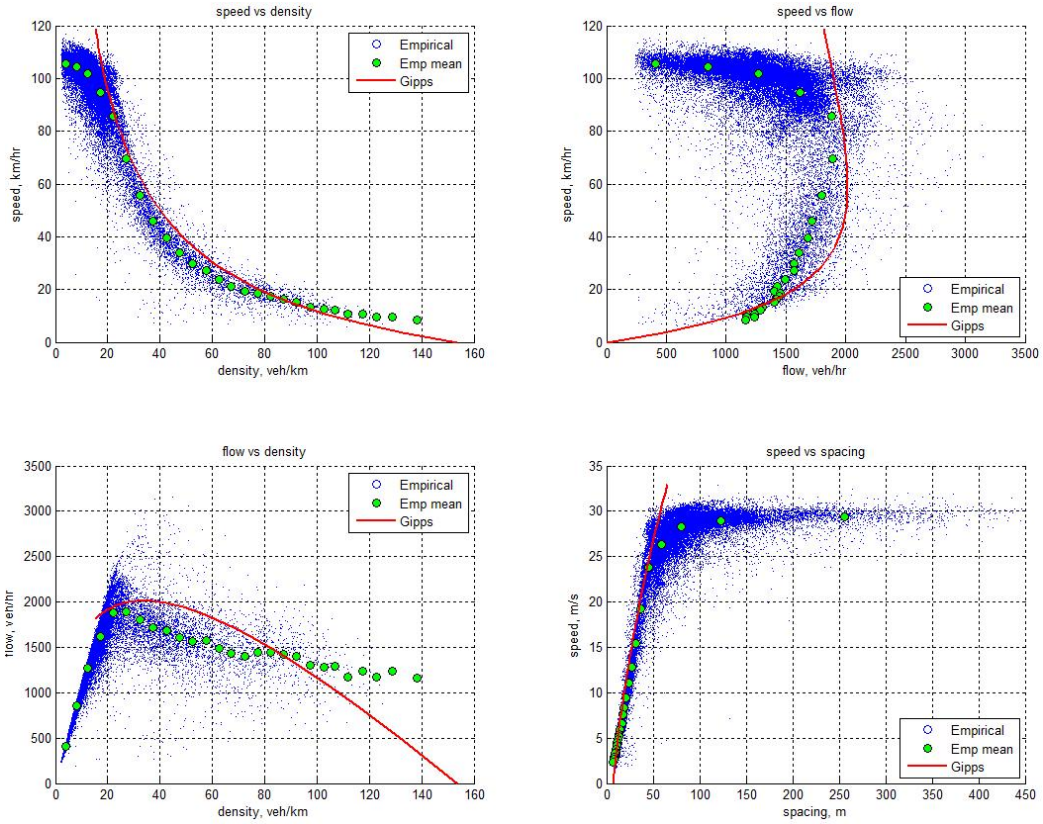


Figure 20.3: Fundamental diagram implied by Gipps model

The above benchmarking is based on the set of parameters in Table 20.2 and the outcome may vary under different set of parameters.

Table 20.2: Macroscopic benchmarking parameters of Gipps model

b	B	τ	l
-3.0 m/s^2	-3.5 m/s^2	1 s	6.5 m

Chapter 21

More Single-Regime Models

This chapter will present a few more car-following models including Newell nonlinear model, Newell simplified model, Intelligent Driver Model (IDM), and Van Aerde Model. Except for Newell simplified model, these models not only capture the essence of car-following behavior but also aggregate to sound macroscopic behavior.

21.1 Newell Nonlinear Model

Newell actually proposed two car-following models, one in 1961 which will be referred to as "Newell nonlinear" car-following model [99] and the other in 2002 which will be referred to as "Newell simplified" car-following model thereafter [103]. The nonlinear car-following model takes the following form:

$$\dot{x}_i(t+\tau_i) = v_i(1 - e^{-\frac{\lambda_i}{v_i}(s_i(t)-l_i)}) \tag{21.1}$$

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 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 j, and l_i is the minimum value of s_i which can be viewed as the nominal vehicle length.

21.1.1 Properties of 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 two thing:

- Unlike Pipes, Forbes, and Gipps models which are derived from driving experiences such as safety rules, this model seems to have no basis on driving experiences but rather a discovery after some contemplation and empirical studies.
- 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 condition, Equation 21.1 reduces to the following speed-density relationship:

$$v = v_f(1 - e^{-\frac{\lambda}{v_f}(\frac{1}{k} - \frac{1}{k_j})}) \tag{21.2}$$

21.1.2 Benchmarking

Microscopic benchmarking refers to the scenario presented in 16.3.1 and macroscopic benchmarking refers to the scenario presented in 16.3.2.

Microscopic Benchmarking

Benchmarking result of Newell Non-linear model is plotted in Figure 21.1. The performance of Newell Non-linear model is summarized as follows:

- Start-up: the model is able to start a vehicle up from stand-still. See the figure when $t > 0$ s.
- Speed-up: 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 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, Newell non-linear model model ceases to be a steady-state model and, instead, becomes a dynamic model. See the figure when $0 < t < 100$ s.

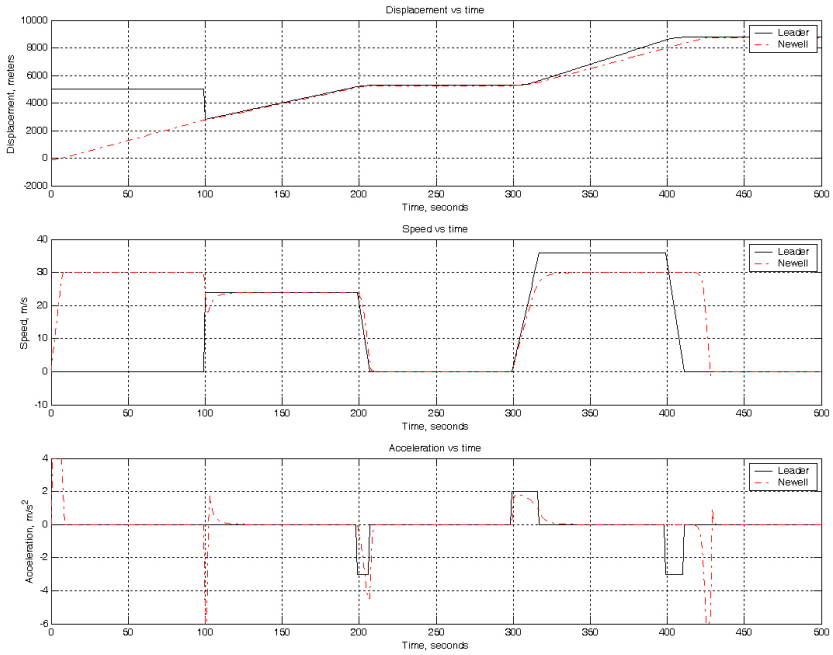


Figure 21.1: Microscopic benchmarking of Newell Non-linear model

- Free-flow: the model is able to reach and settle at desired speed under free-flow condition. See the figure when $0 < t < 100$ s
- Cut-off: By itself, Newell non-linear 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 maximum acceleration and deceleration. Hence, the same argument as that in speed-up applies here. See the figure around $t = 100$ s after incorporating these external conditions.
- Following: the model is able to adopt the leader's speed and follow the leader by a reasonable distance. See the figure when $100 < t < 200$ s.
- Stop-and-go: the model is able to stop the vehicle safely behind its leader and start moving when the leader departs. See the figure when $200 \geq t \leq 300$ s.
- Trailing: the model is able to speed up normally without being tempted by its speeding leader. See the figure when $300 < t < 400$ s.

- Approaching: with the above external logic on limiting deceleration, the model is able to decelerate properly when approaching a stationary vehicle. See the figure when $400 \geq t < 420$ s.
- Stopping: the model is able to stop the vehicle safely behind the stationary vehicle. See the figure when $t \geq 420$ s.

The above benchmarking is based on the set of parameters in Table 21.1 and the outcome may vary under different set of parameters.

Table 21.1: Microscopic benchmarking parameters of Newell non-linear 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

Macroscopic Benchmarking

The fundamental diagram implied by Newell nonlinear model is presented in Figure 21.2 where model parameters are adopted from Newell's original paper.

Newell non-linear model indeed exhibits the correct shape that resembles field observations in the entire density range, as claimed by Newell. First, the model meets 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 it employs only three parameters.

The above benchmarking is based on the set of parameters in Table 21.2 and the outcome may vary under different set of parameters.

Table 21.2: Macroscopic benchmarking parameters of Newell nonlinear model

v_f	k_j	λ
29.5 m/s	1/5 veh/m	0.8

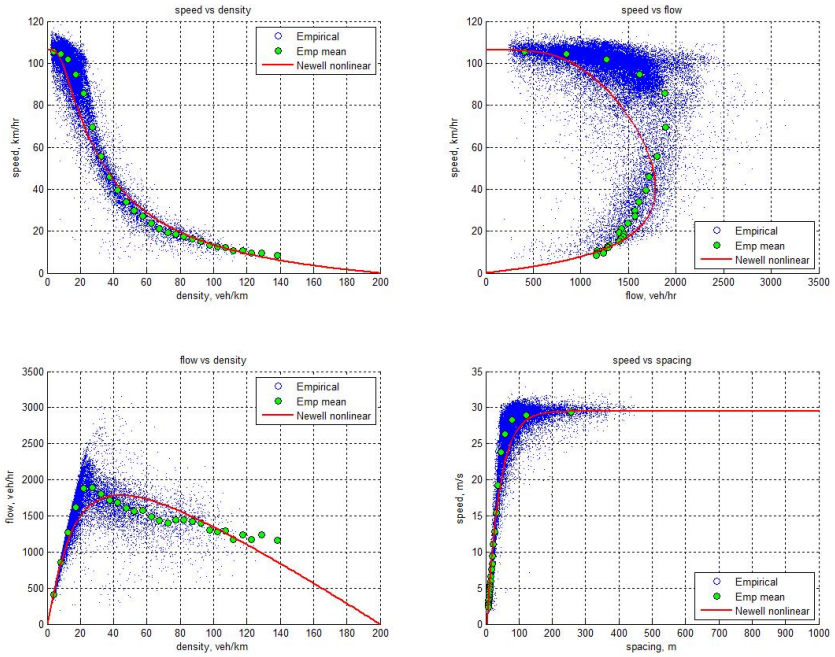


Figure 21.2: Fundamental diagram implied Newell nonlinear model

21.2 Newell Simplified Model

After about 40 years, Newell published a simplified car-following model [103]. 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, assuming that vehicle $i-1$'s trajectory $x_i(t)$ is given in the right pane of Figure 21.3, vehicle i's trajectory can be directly determined by the following equation:

$$x_i(t+\tau_i) = x_{i-1}(t) - l_i \tag{21.3}$$

Graphically, this means to translate 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, i.e., one can squeeze in a rectangle with dimension $\tau_i \times l_i$ between the two trajectories. From the speed-spacing relationship in the right pane, it becomes clear that the physical meaning of l_i is the minimum value of spacing, i.e., the nominal vehicle length, and τ_i is the reciprocal of the tangent to the speed-spacing relationship emitted at point $(0, l_i)$. Evidences show that τ_i

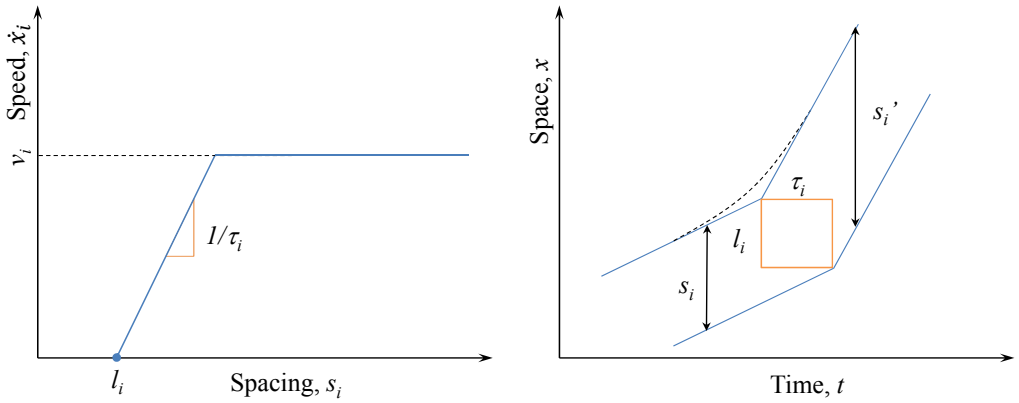


Figure 21.3: Newell simplified car-following model

can most likely be interpreted as the perception-reaction time of driver i. Figure 21.3 also reveals that the spacing between the two vehicles is:

$$s_i(t) = x_{i-1}(t) - x_i(t) \tag{21.4}$$

In addition, the locations of vehicle i at time t and $t + \tau_i$ can be related as:

$$x_i(t + \tau_i) = x_i(t) + \dot{x}_i(t)\tau_i \tag{21.5}$$

Combining the above three equations:

$$s_i(t) = \dot{x}_i(t)\tau_i + l_i \tag{21.6}$$

This is the same as Pipes/Forbes model (by taking the minimum spacing) which, in turn, is equivalent to GM1. Since Newell's simplified model is essentially Pipes/Forbes model, properties and benchmarking of the latter apply to the former.

21.3 IDM Model

The Intelligent Driver Model (IDM) [145, 55] is expressed as a superposition of the follower i's acceleration term and a deceleration term which depends on the desired spacing s_i^*:

$$\ddot{x}_i(t+\tau_i) = g_i[1 - (\frac{\dot{x}_i}{v_i})^\delta - (\frac{s_i^*}{s_i})^2] \tag{21.7}$$

where δ is acceleration exponent, $s_i = x_{i-1} - x_i$ is the spacing between vehicle i and its leader $i-1$, and desired spacing s_i^* is a function of speed $\dot{x}_i$ and relative speed $(\dot{x}_i - \dot{x}_{i-1})$:

$$s_{ij}^* = s_0 + s_1\sqrt{\frac{\dot{x}_i}{v_i}} + T_i\dot{x}_i + \frac{\dot{x}_i[\dot{x}_i - \dot{x}_{i-1}]}{2\sqrt{g_i b_i}} \tag{21.8}$$

where s_0, s_i, and T_i are parameters.

21.3.1 Properties of IDM Model

Under equilibrium condition, Equation 21.7 reduces to the following density-speed relationship:

$$k = \frac{1}{(s_0 + vT)[1 - (\frac{v}{v_f})^\delta]^{-1/2}} \tag{21.9}$$

If one further assumes that $s_0 = s_1 = 0$ and $\delta = 1$, a special cased of Equation 21.9 is resulted:

$$v = \frac{(s-L)^2}{2v_f T^2}[-1 + \sqrt{1 + \frac{4T^2 v_f^2}{(s-L)^2}}] \tag{21.10}$$

where T is average safe time headway and $s = 1/k$ is average spacing and k traffic density.

21.3.2 Benchmarking

Microscopic benchmarking refers to the scenario presented in 16.3.1 and macroscopic benchmarking refers to the scenario presented in 16.3.2.

Microscopic Benchmarking

Benchmarking result of IDM is plotted in Figure 21.1. The performance of IDM is summarized as follows:

- Start-up: the model is able to start the vehicle up from stand-still. See the figure when $t > 0$ s.

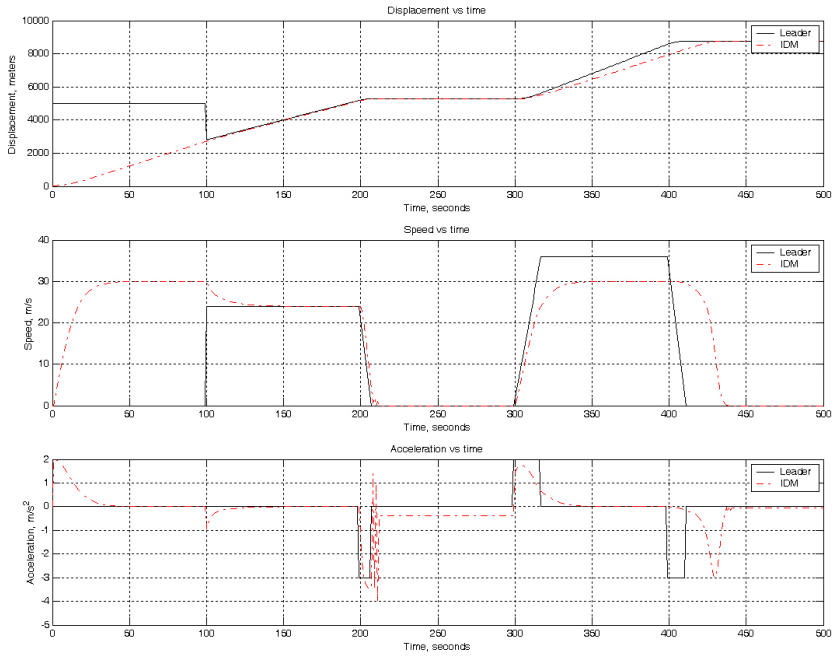


Figure 21.4: Microscopic benchmarking of IDM

- Speed-up: the model is able to speed the vehicle up realistically to its desired speed. See the figure when $0 < t < 100$ s.
- Free-flow: the model is able to reach and settle at desired speed under free-flow condition. See the figure when $0 < t < 100$ s
- Cut-off: the model retains control and responds reasonably when a vehicle cuts in in front. See the figure around $t = 100$ s.
- Following: the model is able to adopt the leader's speed and follow the leader by a reasonable distance. See the figure when $100 < t < 200$ 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 the figure when $200 \geq t \leq 300$ s.
- Trailing: the model is able to speed up normally without being tempted by its speeding leader. See the figure when $300 < t < 400$ s.
- Approaching: the model is able to decelerate properly when approaching a stationary vehicle. See the figure when $400 \geq t < 420$ s.
- Stopping: the model is able to stop the vehicle safely behind the stationary vehicle. See the figure when $t \geq 420$ s.

The above benchmarking is based on the set of parameters in Table 21.3 and the outcome may vary under different set of parameters.

Table 21.3: Microscopic benchmarking parameters of 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

Macroscopic Benchmarking

The fundamental diagram implied by IDM, in particularly Equation 21.9, is presented in Figure 21.5. The model employs four parameters and exhibits a desirable shape with good fitting quality.

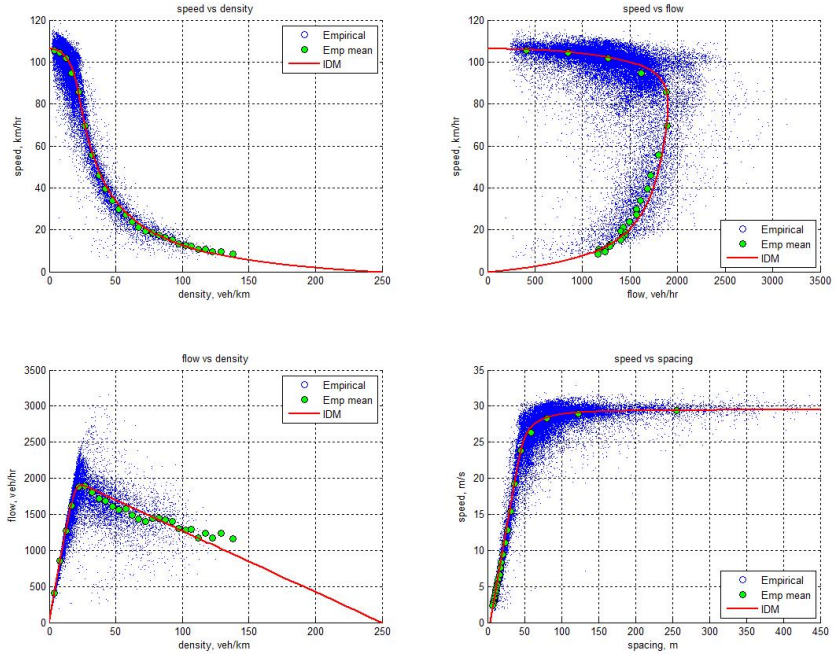


Figure 21.5: Fundamental diagram implied by IDM

The above benchmarking is based on the set of parameters in Table 21.4 and the outcome may vary under different set of parameters.

Table 21.4: Macroscopic benchmarking parameters of IDM

v_f	T	δ	s_0
29.5 m/s	1.7 s	15	4 m

21.4 Van Aerde Model

The Van Aerde car-following model [147, 148] combines Pipes model [119] and Greenshields model [50] into a single equation:

$$s_i = c_1 + c_3\dot{x}_i + c_2/(v_f - \dot{x}_i) \tag{21.11}$$

where

$$\begin{cases} c_1 &= \frac{v_f}{k_j v_m^2}(2v_m - v_f) \\ c_2 &= \frac{v_f}{k_j v_m^2}(v_f - v_m)^2 \\ c_3 &= \frac{1}{q_m} - \frac{v_f}{k_j v_m^2} \end{cases} \tag{21.12}$$

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 occurred at capacity q_m.

21.4.1 Properties of Van Aerde Model

Under equilibrium condiction, Equation 21.11 reduces to the following density-speed relationship:

$$k = \frac{1}{c_1 + c_3 v + c_2/(v_f - v)} \tag{21.13}$$

where all variables are defined before.

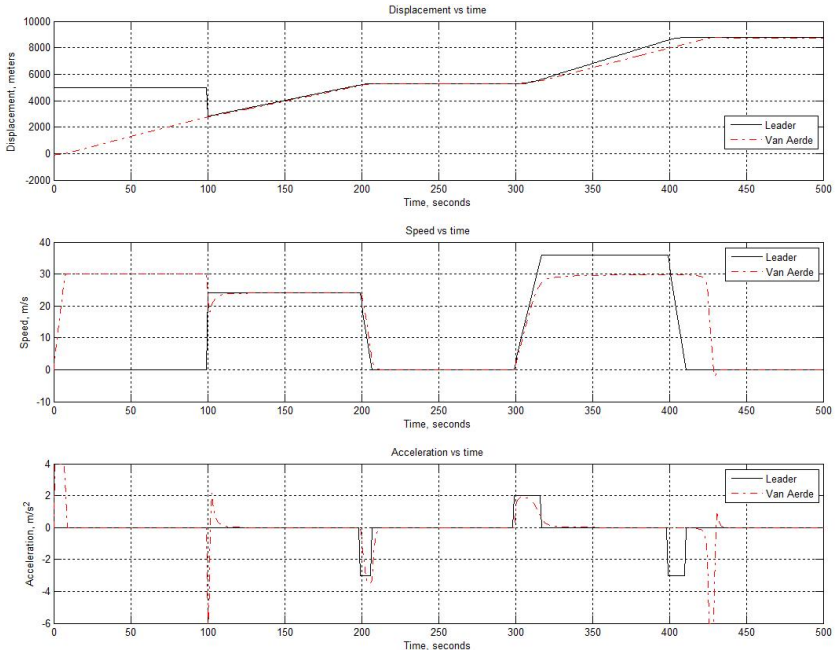


Figure 21.6: Microscopic benchmarking of Van Aerde model

21.4.2 Benchmarking

Microscopic Benchmarking

Benchmarking result of Van Aerde model is plotted in Figure 21.6. The performance of Van Aerde model is summarized as follows:

- Start-up: the model is able to start the vehicle up from stand-still. See the figure when $t > 0$ s.
- Speed-up: the same argument as that in the corresponding part of Newell non-linear car-following model applies here. See the figure when $0 < t < 100$ s.
- Free-flow: the model is able to reach and settle at desired speed under free flow condition. See the figure when $0 < t < 100$ s
- Cut-off: the same argument as that in the corresponding part of Newell non-linear car-following model applies here. See the figure around $t = 100$ s.
- Following: the model is able to adopt the leader's speed and follow the leader by a reasonable distance. See the figure when $100 < t < 200$ s.

- Stop-and-go: the model is able to stop the vehicle safely behind its leader and start moving when the leader departs. See the figure when $200 \geq t \leq 300$ s.
- Trailing: the model is able to speed up normally without being tempted by its speeding leader. See the figure when $300 < t < 400$ s.
- Approaching: with limiting deceleration, the model is able to decelerate properly when approaching a stationary vehicle. See the figure when $400 \geq t < 420$ s.
- Stopping: the model is able to stop the vehicle safely behind the stationary vehicle. See the figure when $t \geq 420$ s.

The above benchmarking is based on the set of parameters in Table 21.5 and the outcome may vary under different set of parameters.

Table 21.5: Microscopic benchmarking parameters of Van Aerde model

k_j	v_f	τ_i	v_m	q_m
1/6 veh/m	30 m/s	1.0 s	25 m/s	1800 veh/h
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	-99.4 m	0 m/s	0 m/s^2

Macroscopic Benchmarking

The fundamental diagram implied by Van Aerde model is presented in Figure 21.7. 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 21.6 and the outcome may vary under different set of parameters.

Table 21.6: Macroscopic benchmarking parameters of Van Aerde model

v_f	k_j	v_m	q_m
29.5 m/s	1/4 veh/m	20 m/s	1950 veh/h

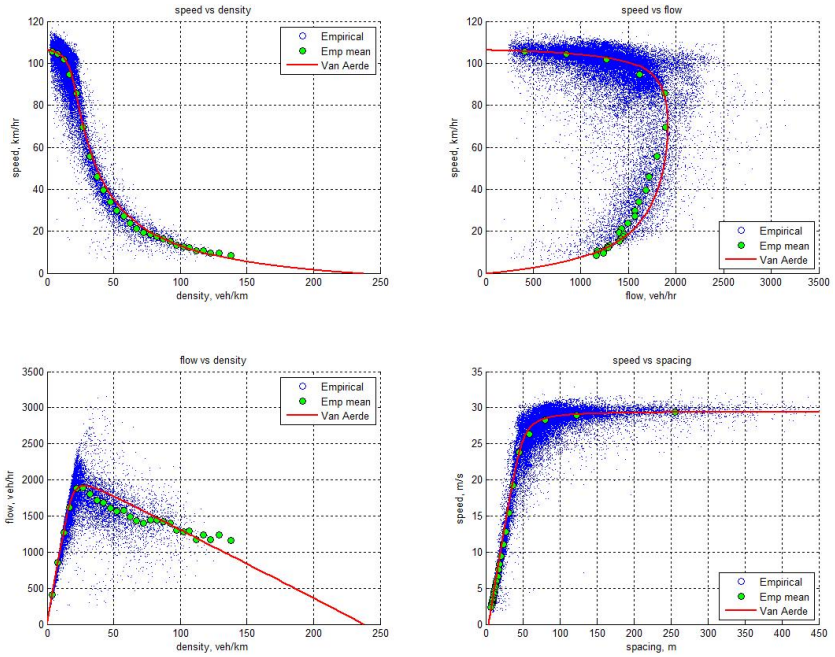


Figure 21.7: Fundamental diagram implied by Van Aerde model

Chapter 22

More Intelligent Models

Car-following models introduced up to this point share one thing in common, i.e., they are one-equation models except for Gipps model which has two equations. This means that these models use a single equation to handle all driving situations including start-up, speed-up, free-flow, approaching, following, and stopping. Hence, these models are referred to as single-regime models. Gipps model is a two-regime model since it has an equation for free-flow and another for car following. A model is called multi-regime if it differentiates driving regimes and handles them using different equations. Car-following models introduced in this chapter fall into this category.

22.1 Psycho-Physical Model

A typical Psycho-Physical model is the one proposed by Wiedemann [157] in 1974. The model considers tow major factors influencing driver's operational control: relative position Δx (i.e. spacing $x_{i-1} - x_i$) and relative speed $\Delta \dot{x}$ (i.e. the opposite of speed difference $\dot{x}_i - \dot{x}_{i-1}$). Hence, the working principle of the model can be illustrated by a diagram using $\Delta \dot{x}$ as the horizontal axis and Δx as the vertical axis, see Figure 28.2.

The operating condition of a vehicle i in relation to its leading vehicle $i-1$ can be represented as a point ($\Delta \dot{x}$, Δx) in the diagram. As vehicle i moves (also in relation to $i-1$), its operating point changes accordingly, leaving a trajectory in the diagram. The relation of the two vehicles can be interpreted by examining the location of the operating point. For example, if the point is in the the negative side of $\Delta \dot{x}$, vehicle i is traveling slower

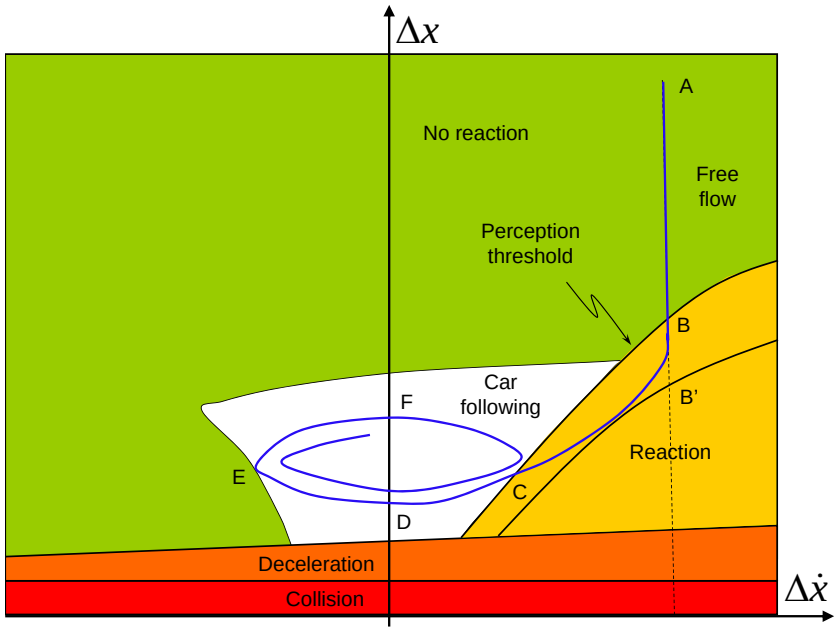


Figure 22.1: Illustration of a psycho-physical model

than vehicle $i-1$, while the relation reverses if the point is in the positive side of $\Delta\dot{x}$. In addition, the point is always in the positive side of Δx since vehicle $i-1$ takes the lead. The smaller Δx is, the closer the two vehicles are. Hence, the two vehicles collide if Δx is less than one vehicle length l. This 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 as 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 $i-1$. 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 doesn't have to respond to $i-1$, an area of which is denoted as *no reaction* in the diagram. As vehicle i keeps moving, relative speed $\Delta\dot{x}$ remains unchanged, but relative position 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 cut-off 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 point B and B' vary as $\Delta\dot{x}$ changes. The trajectory of B or B' under different $\Delta\dot{x}$ separates *reaction* are from *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 car following during which driver i tries to keep the same pace as $i-1$ 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. using cell phone), during which $\Delta\dot{x}$ becomes negative and keeps decreasing while Δx increase. 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 $\Delta\dot{x}$ 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 to G. Therefore, as the driver oscillates back and forth around his her comfortable car-following distance, the operating point drifts around within an area in the diagram denoted as *car following*.

The model derived its name as a psycho-physical model 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 models introduced before, this model captures more driving regimes such as free-flow (*no reaction* area), approaching (*reaction* area), following (*car following* area), and decelerating (*deceleration* area).

22.2 CARSIM Model

The CARSIM model [9] is another multi-regime 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 i's initial speed and urgency of the task, the acceleration rate is

found by entering Tables 1 and 2 provided in the original paper.

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.

A3: Vehicle i was stopped and has to start from stand-still. A maximum acceleration rate is applied constrained by a non-collision constraint after a response delay.

A4: Vehicle i is in car-following with its leader $i-1$. A4 is determined by satisfying the following safety rule: vehicle i should leave a non-negative gap ($s_i - l_{i-1} \geq 0$) from $i-1$ should i be advanced one time increment 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 other variables are as defined before.

A5: Vehicle i in car-following is subject to a non-collision constraint which is reinforced by considering the desired spacing:

$$s_i^*(t) = x_{i-1}(t) - x_i(t+\Delta t) \geq max \begin{cases} \dot{x}_i(t+\Delta t)\tau_i + l_{i-1} \text{ or} \\ \dot{x}_i(t+\Delta t)\tau_i + \frac{[\dot{x}_i(t+\Delta t)]^2}{(2B_i)} - \frac{[\dot{x}_{i-1}(t)]^2}{(2B_{i-1})} + l_{i-1} \end{cases}$$

where $\dot{x}_i(t+\Delta t) = \dot{x}_i(t) + A_5 \Delta t$, B_i and B_{i-1} are maximum deceleration rate of i and $i-1$ respectively. The astute reader immediately recognizes that the first choice of the right-hand side follows the rationale of Forbes model [39, 38, 37] and the second choice is similar to that of Gipps [46] model if driver i is willing to apply emergency brake (i.e. $b_i = B_i$) as well.

22.3 Rule-Based Model

A representative of rule-based models is the one developed by Kosonen [76] which is reproduced below:

1. NO SPEED CHANGE

   Keep the present speed level (default case).

2. ACCELERATE IF [$\dot{x}_i < v_i$] and [$t - t_{last} > T_{acc}(\dot{x}_i)$]

   The current speed $\dot{x}_i$ is less than desired speed v_i and the time elapsed from last acceleration t_{last} is more than T_{acc}.

3. NO ACCELERATION IF [$s_{ij} < s_{min}(\dot{x}_i, \dot{x}_j) + w_{stab}(\dot{x}_i, \dot{x}_j)$]

   The distance from obstacle s_{ij} is less than the minimum safe distance s_{min} plus the width of stable area w_{stab}.

4. SLOW DOWN IF [$s_{ij} < s_{min}(\dot{x}_i, \dot{x}_j)$]

   The distance from obstacle Dobs is less than the minimum safe distance s_{min}.

5. DO NOT SLOW DOWN IF [$\dot{x}_i < \dot{x}_j$] or [$t - t_{last} < T_{maxdec}$]

   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).

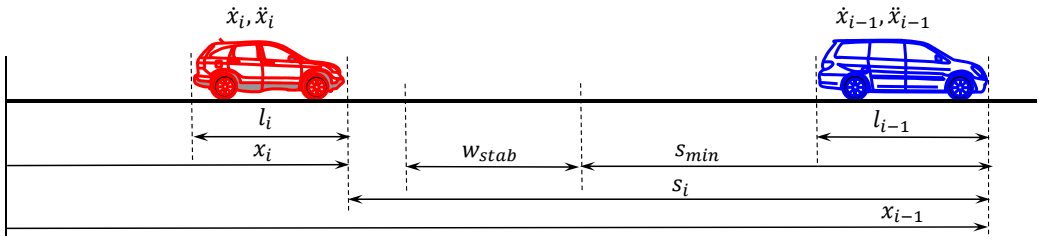


Figure 22.2: Illustration of a rule-based model

At each time step, the motion of vehicle i is checked against the above rules one by one. A latter rule always supersedes earlier ones should there be a conflict. Compared with models presented before, the rule-based mode is even closer to human intelligence with less mathematical tractability.

22.4 Neural Network Model

Perhaps the approach that best mimics driver behavior is artificial neural networks (ANNs) [68] [112]. This is because ANNs are capable of making association, recognizing patterns, learning and memorizing experiences, self-organizing, and adapting to new environment. A neural network typically

consists of many interconnected working units called neurons, see Figure 22.3 for an example of neural network in the right pane and a neuron in the left pane. A neuron receives inputs x_1, x_2,..., x_n which are weighted w_1, w_2,..., 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 the threshold of the neuron θ. The neuron outputs 1 if $z \geq \theta$ and 0 otherwise.

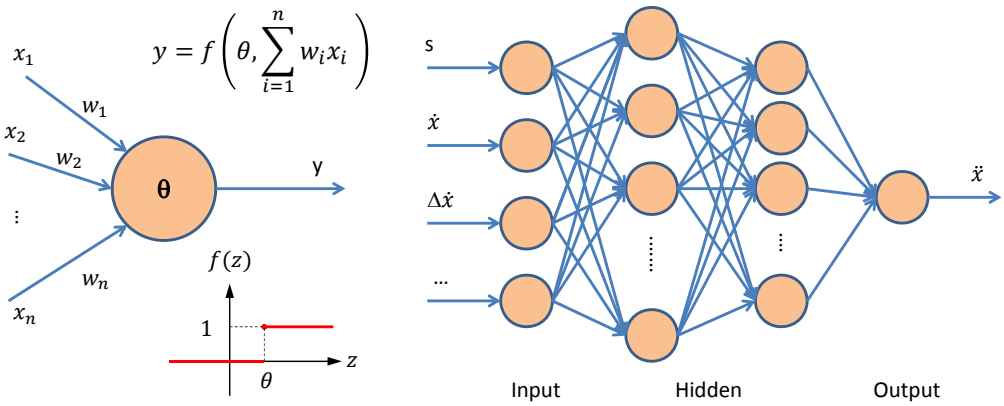


Figure 22.3: Illustration of a neural network model

Neurons with such a simple functionality can be organized into neural networks of varying complexity and topology. The right pane of Figure 22.3 illustrates an example of back-propagation (BP) 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 only feeds its output forward to neurons in the next layer without backward feeding and cross-layer connection.

To apply neural network to modeling car-following behavior, one first identifies a set of factors to be considered that influence 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, inter-vehicular communication, etc. 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 lies 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 is 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 $[s(1), \dot{x}(1), \Delta\dot{x}(1), ...]$ results in driver operational control $[\ddot{x}(1)]$, and more patterns are observed at $t_2, t_3, ..., t_m$:

$$\begin{bmatrix} s(1), \dot{x}(1), \Delta\dot{x}(1), ... \\ s(2), \dot{x}(2), \Delta\dot{x}(2), ... \\ ... \\ s(m), \dot{x}(m), \Delta\dot{x}(m), ... \end{bmatrix} \Rightarrow \begin{bmatrix} \ddot{x}(1) \\ \ddot{x}(2) \\ ... \\ \ddot{x}(m) \end{bmatrix} \tag{22.1}$$

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) to the input layer which feeds forward to hidden layers and eventually to the output layer. If the computed output is different than the observed output, the error is propagated backwards layer by layer to adjust their connection weights and neuron thresholds. This is why networks of this kind are called back-propagation. 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 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 till 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 22.1). This is followed by 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 to the corresponding input.

The trained neural network is now ready to apply 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) based on 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 new environment which the network has never encountered before.

Chapter 23

Summary of Car-Following Models

It is time to summarize car-following models introduced so far. One way to classify these models is to examine their model intelligence. For example, Pipes model is a one-equation model and this equation handles all driving situations, i.e., they are treated as a single regime. Hence Pipes model is a single-regime model. Also in this category are Forbes model, GM models, Newell models, IDM, and Van Aerde model. Gipps model consists of two equations, one for free-flow and the other for car following, and hence is a two-regime model. Both Carsim model and Psycho-Physical model differentiate more driving regimes and hence are multi-regime models. Further, rule-based model incorporates driving strategies for various driving regimes into a set of simple IF-THEN rules. Better yet, neural network model applies artificial intelligence to learn, memorize, and adapt to driving experiences. Illustrated in Figure 23.1, car-following models become more and more intelligent as one moves from left to right.

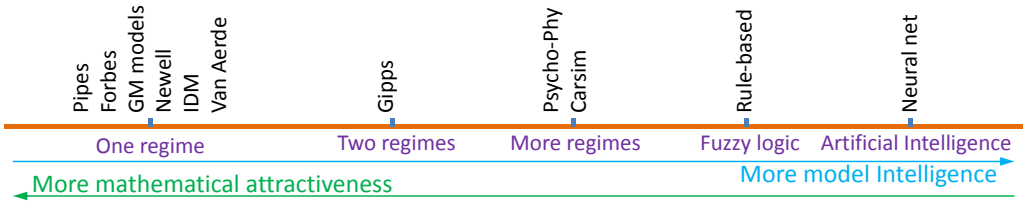


Figure 23.1: Summary of car-following models

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 mode in order to understand its macroscopic properties. Therefore, single-regime model are mathematically attractive. Two- or multi-regime models, however, are inevitably piece-wise and involves discontinuity which makes them less mathematically attractive. Though computationally simple, rule-based model consists of a set of IF-THEN rules rather than clearly defined mathematical formulation. Hance, it is very difficult to analyze macroscopic properties of this kind of model. Neural network model, to the extreme, is very intractable because there is no clear, mathematical formulation that defines the relation between input variables and output variable. If a model with clear mathematical formulation is like a crystal box through which one can trace an input all the way to the output, a neural network is like a black box through which one knows which input leads to which output but what happens in between is a mystery.

A more rigorous effort of the taxonomy of microscopic models is made by the Next-Generation Simulation (NGSIM) program[1], see the diagram in Figure 23.2. 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 car-following module, there are a few lines representing different modeling approaches. For example, one approach is called *Stimulus-Response* which starts with with a few papers published around 1960 serving as the basis of GM models. Labeled along this line are further models proposed or existing models revised, showing historical evolution of this modeling approach. One line up is *Desired Measure* approach, along which are Pipes model, Newell non-linear model, Gipps model, and Carsim model. The next line is *Psycho-Physical* approach where one finds Wiedeman model. This is followed by the *Rule-Based* approach, an example of which is Kosonen model. IDM is on its own at the top. Note that a potential addition to this module can be one more approach called Neural Network. Other modules and models shown therein can be interpreted in a similar way.

On the right-hand side of the diagram are a set of vertical lines. On top of them are a set of transportation simulators (or simulation software packages), among which are AIMSUN, CORSIM, HUTSIM, Integration, Paramics, VISSIM, etc. The intersection of a horizontal line (a modeling approach) and a

[1]http://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm

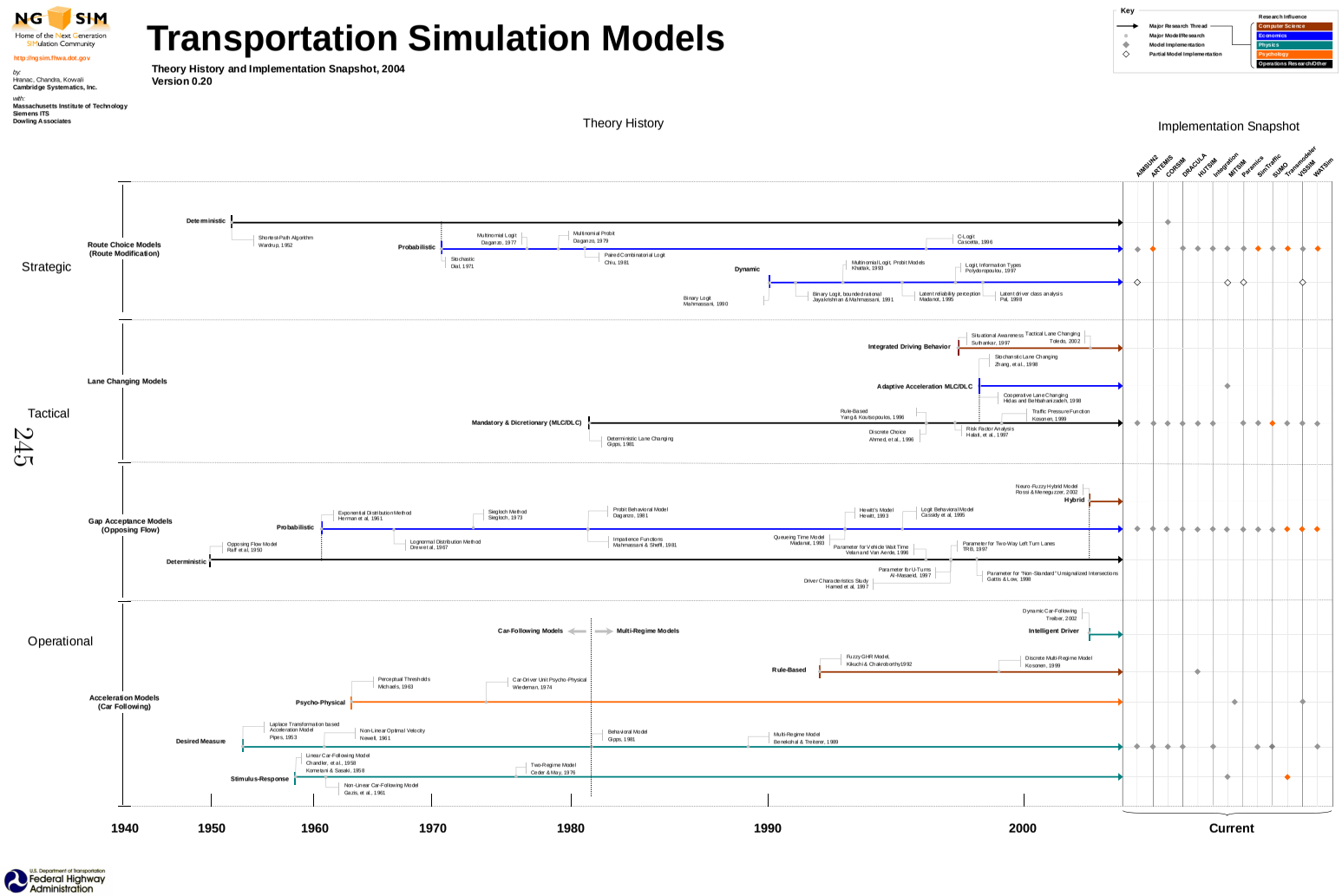


Figure 23.2: Taxonomy of microscopic models

vertical line (a simulator) denote potential implementation of a car-following model of this approach in the simulator. If the implementation is true, a diamond-shaped dot is placed on the intersection. Therefore, it is clear that car following in CORSIM is handled by a *Desired Measure* model, car following in VISSIM is handled by a *Psycho-Physical* model, and car following in HUTSIM is handled by a *Rule-Based* model. The association of simulators and models in other modules can be interpreted in a similar way.

Part IV

Picoscopic Modeling

Chapter 24

Picoscopic Modeling

Suppose one is observing traffic 10,000 m above the ground, traffic behaves as a compressible fluid whose states (speed, flow, and density, etc.) propagate back and forth like waves. This is a scenario of *macroscopic* modeling. If one lowers to 3,000 m, 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 lowers even more to 1,000 m, 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.

Figure 24.1: A picoscopic view of transportation system

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 wave nor particle, but a detailed picture incorporating drivers, vehicles, and environment (e.g. roadway, signs, signals, etc.), see Figure 24.1. 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.

24.1 Driver, Vehicle, and Environment

Traffic flow modeling at the picoscopic level should not only represent drivers, vehicles, and environment in different models but 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 [105, 108] illustrated in Figure 24.2

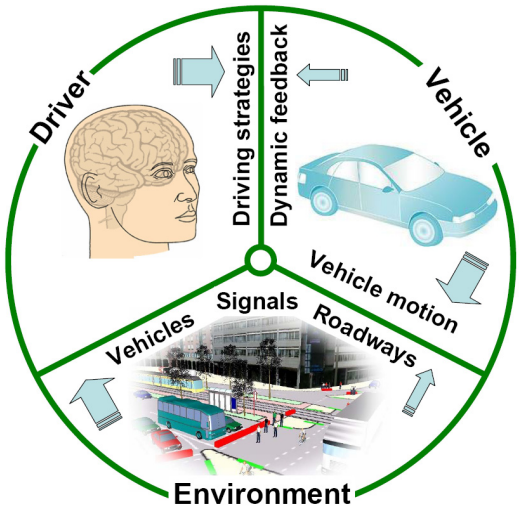


Figure 24.2: A driver-vehicle-environment closed-loop system

In a transportation system, drivers are active components which make

decisions, while vehicles are passive components which execute decisions. The interaction between a driver and his/her vehicle constitutes a basic unit in 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 based on their knowledge. Figure 24.3 presents the structure of such a driver modeling approach.

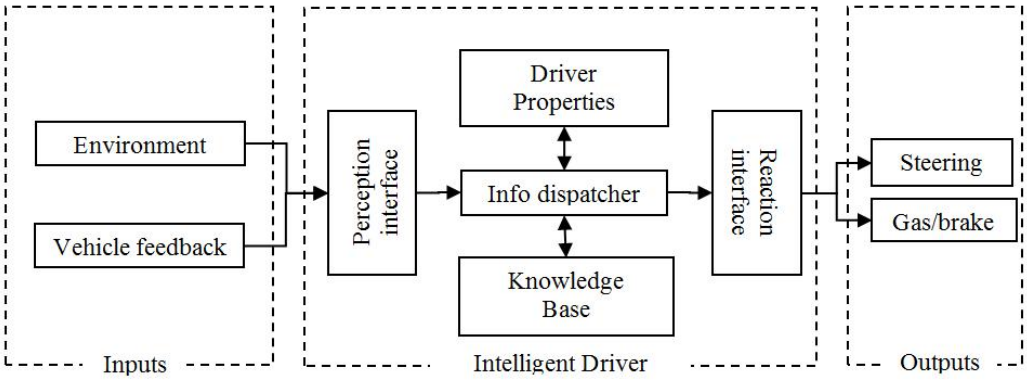


Figure 24.3: Picoscopic modeling: driver modeling

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 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) exercises 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 based on 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 interprets driver decisions to actionable instructions before it is 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, gas, and brake.

In Chapter 27, 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 24.4 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 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.

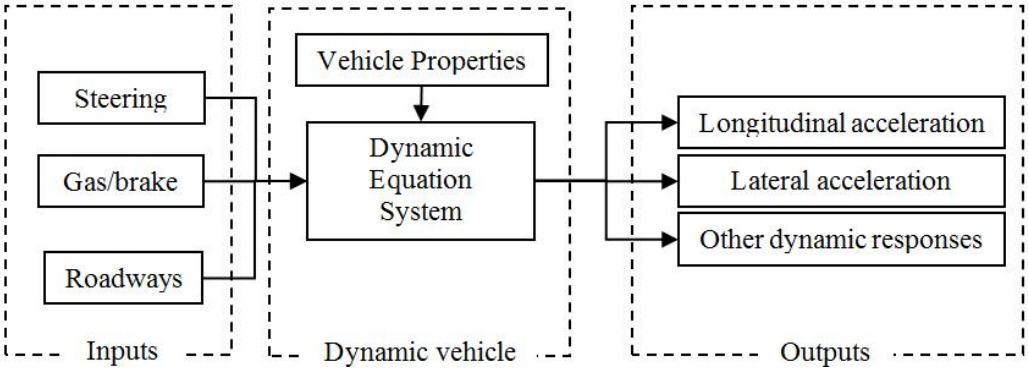


Figure 24.4: Picoscopic modeling: vehicle modeling

In Chapter 25, a simple engine model will be formulated with reasonable accuracy and excellent computational efficiency to facilitate vehicle modeling. In Chapter 26, a simple Dynamic Interactive Vehicle (DIV) model will be formulated that requires little calibration effort and fewer computational resources.

Combining the above driver and vehicle models forms a driver-vehicle unit which constitutes a basic building block of roadway traffic. Many such units as well as roadways, traffic control devices, and other transportation system components constitute a general environment in which the driver-vehicle unit operates. The interactions among drivers, vehicles, and environment are summarized in the picoscopic transportation modeling architecture shown in Figure 24.5.

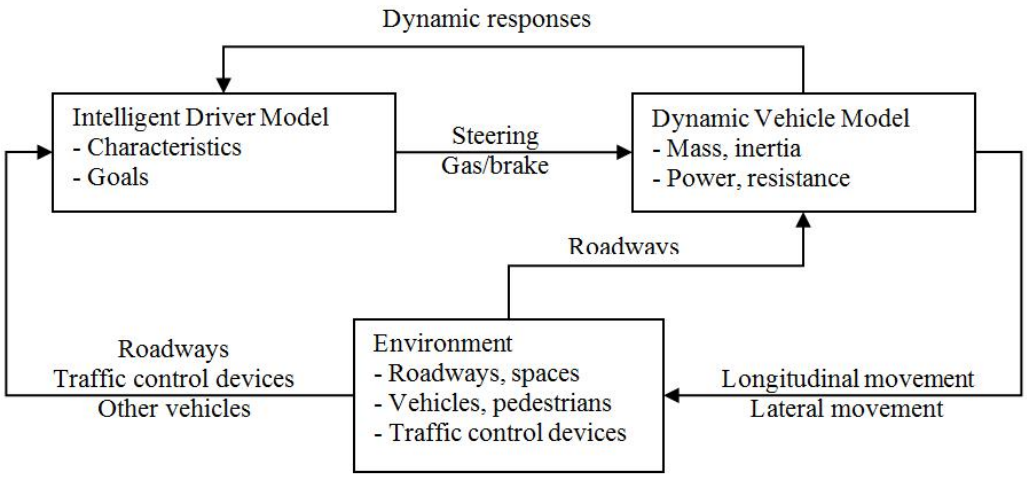


Figure 24.5: Picoscopic modeling: modeling architecture

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 gas/brake). 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 based on vehicle dynamic equations. Moving longitudinally and laterally, the vehicle constitute part of the environment. Part of vehicle dynamic responses such as speed, acceleration, and yaw velocity are fed back to the driver for determining 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 movements and interactions of all vehicles in the environment.

24.2 Applications of Picoscopic Modeling

Transportation modeling and simulation are 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). Considering the processing power of today's computers, a macroscopic model can achieve 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 ultra-high modeling resolution but within a very limited geographical area such as the roads surrounding Public Garden. With such a fine level of detail, picoscopic modeling can help address many transportation-related problems, among which the following are a few examples.

24.2.1 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 computer to check if the highway provides sufficient sight distance to avoid accidents or a curve 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.

24.2.2 Connected Vehicle Technology

Future vehicles will be equipped with Dedicated Short Range Communications (DSRC) 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 an illustration in Figure 24.6. 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 incorporating the effects of connected vehicle technology into modeling.

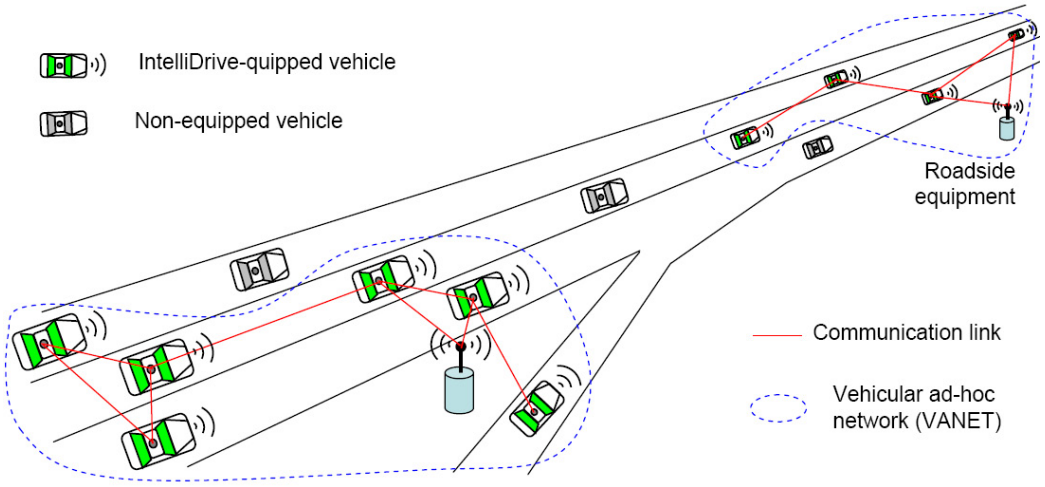


Figure 24.6: An illustration of connected vehicle technology

24.2.3 Transportation Forensics

Investigation of a traffic accident frequently requires the ability to decipher what happens shortly before, during, and after the accident. This involves reconstructing the accident during which the driver perceives immediate hazard, makes decision, and executes control, while the vehicles 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.

24.2.4 Emergency Management

In analyzing transportation systems under extreme conditions, it is essential to have both capabilities of overseeing the full picture (e.g. a regional transportation network) and 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. 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 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 usual. As a result, safety as a primary goal may give its 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 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" which prevents these tools from modeling and simulating transportation systems under extreme conditions.

Chapter 25

Engine Modeling

This chapter is a reproduction of the work in reference [109].

25.1 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 potential collisions. In order to ensure safety, this procedure has to be repeated with such a high frequency that conventional engine models, due to their intrinsic complexity, are beyond the capacity of a contemporary on-board computer. In addition, most of these engine models require proprietary parameters such as throttle body size and mass of piston. This prevents the adoption of these models across a wide variety of vehicle platforms. Therefore, special criteria apply to an ideal engine model suited for the above-mentioned applications. Most salient of these criteria are the following:

- 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 on-board computing with high frequency in real time.
- Accessibility: to assist wide deployment across different vehicle platforms, 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 their associated engine speeds) should be publicly available (e.g. http://www.cars.com).
- Formulation: the engine model should be analytical. Engine models based on look-up tables are not only prohibitive to prepare wide classes of vehicles but also resource-demanding in computation.
- Calibration: The engine model should involve the least calibration effort - calibration-free is most desirable. Again, it would be a daunting task if an engine model has 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 comparing the performance of these models, based on which, the best model will be recommended. Comparing 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's principle) have already been explored in the past. However, the recommended model does fill a gap in non-conventional arena such as CVT-enabled applications where an excellent computational efficiency and reasonable accuracy are desirable.

25.2 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. Nevertheless, the review should present a reasonable overview of historical efforts and the current state.

It appears that efforts of engine modeling have a much shorter history than the engine itself. The first physically based dynamic engine models

were reported in [121], [27], [28], and [15] which recognized the effects of throttle and intake manifold dynamics. Much of the early efforts in engine modeling has been surveyed in [122] with a focus on IC engine models for control.

A trend of increasing modeling accuracy was quite noticeable in the historical evolution. Engine models developed in [96] and [161] included fuel film dynamics and engine rotational dynamics with transport delays. Continuing this modeling approach, A three state engine model was developed in [14] based on the work in [27] [28] and [95]. Shortly after, Akinci et al [5] also presented a nonlinear three state dynamic model of a spark ignition (SI) engine and further effort was reported in [125]. Rizzoni [131] formulated a global model for the IC engine, and a concurrent paper [132] described a stochastic model. A nonlinear engine model was proposed in [17]. Hong [63] developed an engine model based on the "filling and emptying" method for unsteady gas flow across the engine cylinder [61]. A low-dimensional, physically motivated engine model was proposed in [22]. Shiao et al [135] 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, [13] described the unsteady gas flow in both intake and exhaust systems. A very complicated engine model involving twelve degrees of freedom [83] was proposed to capture even more details.

At the applied side, efforts were identified which adopted existing models or extended existing work. [69] applied the two-state engine model developed in [14] to the slip control. A real-time engine model [150] similar to [96] was used to develop a nonlinear model-based control strategy for hybrid vehicles. Delprat et al [26] modeled an IC engine as part of hybrid vehicle modeling. Scillieri et al [134] developed a direct-injection spark-ignition (DISI) engine model to demonstrate the potential performance benefits of reference feedforward control. Two simulation packages involving IC engine models [10] [40] 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 internal combustion engines, Cook and Powell [16] argued that a linear engine model reduced from [121] might be desirable for the purpose of engine control analysis. To facilitate the development of AICC (Autonomous Intelligent Cruise Control), Swaroop et al [141] used an engine model which was essentially the first state equation developed in [14]. A very simple engine model was presented in [51]

for teaching purpose. An even simpler model was suggested by Genta [44] 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 A.

25.3 Simple Mathematical Engine Models

This section presents three simple engine models. Model I is an existing one [44], while Model II and III are developed by the authors.

25.3.1 Model I: Polynomial Model

In an effort to develop a dynamic vehicle model, Genta [44] suggested a very simple engine model which used a polynomial to empirically approximate the relationship between engine power, P, and engine speed, ω, i.e.

$$P = \sum_{i=1}^{3} C_i \omega^i \tag{25.1}$$

where C_i's ($i = 0, 1, 2, 3$) are coefficients and can be estimated from empirical engine curves. Artamonov et al [6] suggested the following values for an SI engine:

$$\begin{array}{rcl} C_1 & = & P_{max}/\omega_p \\ C_2 & = & P_{max}/\omega_p^2 \\ C_3 & = & -P_{max}/\omega_p^3 \end{array} \tag{25.2}$$

where P_{max} is the peak power and ω_p is the engine speed at which power peaks. As is well known, engine torque, Γ, is engine power divided by engine speed:

$$\Gamma = \sum_{i=1}^{3} C_i \omega^{i-1} \tag{25.3}$$

where coefficients C_i ($i = 0, 1, 2, 3$) remain the same as in Equation 25.1.

25.3.2 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:

$$\Gamma = C_1 + C_2(\omega - \omega_t)^2 \tag{25.4}$$

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:

$$P = C_3\omega + C_2(\omega - \omega_t)^2\omega \tag{25.5}$$

Given that the engine outputs P_{max} at ω_p and outputs Γ_{max} at ω_t, the following are resulted:

$$\Gamma_{max} = C_1 + C_2(\omega_t - \omega_t)^2 = C_1 \tag{25.6}$$

$$P_{max} = C_3\omega_p + C_2(\omega_p - \omega_t)^2\omega_p \tag{25.7}$$

$$\frac{dP}{d\omega}|_{\omega=\omega_p} = (C_3\omega + C_2(\omega - \omega_t)^2 + 2C_2\omega(\omega - \omega_t))|_{\omega=\omega_p} = 0 \tag{25.8}$$

Solve equations 25.7 and 25.8:

$$C_2 = \frac{P_{max}}{2\omega_p(\omega_p - \omega_t)} \tag{25.9}$$

$$C_3 = \frac{P_{max}}{2\omega_p}(3\omega_p - \omega_t) \tag{25.10}$$

Therefore

$$\Gamma = \Gamma_{max} + \frac{P_{max}}{2\omega_p(\omega_p - \omega_t)}(\omega - \omega_t)^2 \tag{25.11}$$

$$P = \frac{P_{max}}{2\omega_p}(3\omega_p - \omega_t)\omega + \frac{P_{max}}{2\omega_p(\omega_p - \omega_t)}(\omega - \omega_t)^2\omega \tag{25.12}$$

Equations 25.11 and 25.12 constitutes Model II which guarantees that its power and torque curves peak at their respective peak engine speeds.

25.3.3 Model III: Bernoulli Model

This model is based on Bernoulli's 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 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 preparation, the Bernoulli engine model is developed as follows.

Theoretical volumetric fresh mixture flow rate into the engine, $\dot{V}_t$, is:

$$\dot{V}_t \text{ (m}^3\text{/s)} = V_e \text{ (m}^3\text{/cycle)} \times \text{cycles/rev} \times \text{engine speed (rev/s)} \quad (25.13)$$

where V_e is engine displacement, cycles/rev = 1/2 for a four-stroke engine, and engine speed (rev/s) = ω_e / 2 where ω_e is engine speed in rad/s. Therefore,

$$\dot{V}_t = V_e \times \frac{1}{2} \times \frac{\omega}{2\pi} = \frac{V_e \omega_e}{4\pi} \quad (25.14)$$

This model assumes that the air is an ideal gas. According to ideal gas law:

$$pV = \frac{m}{m'} RT \quad (25.15)$$

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 Kelvin. Therefore, $m = \frac{pm'V}{RT}$ and the density of the gas in the vessel is

$$\rho = \frac{m}{V} = \frac{pm'}{RT} = \frac{p}{R_a T} \quad (25.16)$$

where $R_a = R/m'$ and for air $R_a \approx 287$ Nm/kg/K. Further, the mass air flow rate, $\dot{m}$, as a function of the volumetric air flow rate, $\dot{V}$, is

$$\dot{m} = \frac{pm'}{RT}\dot{V} = \frac{p}{R_aT}\dot{V} \tag{25.17}$$

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 section area of any point in the intake manifold. The constriction in the manifold is the throttle whose cross section area is $\theta \times A$ where θ is percent of throttle opening. So the mass flow rate of air entering the engine is

$$\dot{m} = \frac{p}{R_aT}\dot{V}_t = \frac{p}{R_aT}vA \tag{25.18}$$

According to compressible fluid mechanics [7], the speed of air flow, v, is related to a Mach number, M_a, which is the ratio of air flow speed to sound speed $v_s = \sqrt{kR_aT}$, i.e.

$$M_a = \frac{v}{v_s} = \frac{v}{\sqrt{kR_aT}} = \frac{\dot{V}_t}{A\sqrt{kR_aT}} \tag{25.19}$$

where k is specific heat ratio. Assume stagnation state (where the flow is brought into a complete motionless condition in isentropic process without other forces) holds. Follow the stagnation state for ideal gas model in Sections 4.1 and 4.2 of [7], equation 25.18 can be translated to:

$$\dot{m} = A(\frac{\sqrt{k}M_ap_0}{\sqrt{R_a}T_0})(1 + \frac{k-1}{2}M_a^2)^{-\frac{k+1}{2(k-1)}} \tag{25.20}$$

where p_0 and T_0 are stagnation pressure and temperature, respectively. Plug 25.19 into 25.20 yields

$$\dot{m} = A(\frac{\dot{V}_tp_0}{AR_aT_0})(1 + \frac{\dot{V}_t^2(k-1)}{2A^2kR_aT_0})^{-\frac{k+1}{2(k-1)}} \tag{25.21}$$

Notice that equations 25.20 and 25.21 apply to flow everywhere. When the flow is choked (i.e. $M_a = 1$) and stagnation conditions (i.e. temperature, pressure) do not change, equation 25.20 reduces to:

$$\dot{m} = A(\frac{\sqrt{k}p_0}{\sqrt{R_a}T_0})(1 + \frac{k-1}{2})^{-\frac{k+1}{2(k-1)}} \tag{25.22}$$

Assume exact Stoichiometric air-fuel ratio λ, fuel energy density E_f, engine thermal efficiency η, the power developed by the engine is:

$$P = \lambda E_f \eta [A(\frac{\dot{V}_t p_0}{AR_a T_0})(1 + \frac{\dot{V}_t^2 (k-1)}{2A^2 k R_a T_0})^{-\frac{k+1}{2(k-1)}}] \tag{25.23}$$

Plug in equation 25.14,

$$P = \lambda E_f \eta [A(\frac{V_e \omega_e p_0}{4\pi AR_a T_0})(1 + \frac{V_e^2 \omega_e^2 (k-1)}{32\pi^2 A^2 k R_a T_0})^{-\frac{k+1}{2(k-1)}}] \tag{25.24}$$

The torque that the engine develops is

$$\Gamma = \lambda E_f \eta [A(\frac{V_e p_0}{4\pi AR_a T_0})(1 + \frac{V_e^2 \omega_e^2 (k-1)}{32\pi^2 A^2 k R_a T_0})^{-\frac{k+1}{2(k-1)}}] \tag{25.25}$$

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 empirical engine curves after peak torque and power. Therefore, the engine model is modified by adding a correction term:

$$P = \lambda E_f \eta [A(\frac{V_e \omega_e p_0}{4\pi AR_a T_0})(1 + \frac{V_e^2 \omega_e^2 (k-1)}{32\pi^2 A^2 k R_a T_0})^{-\frac{k+1}{2(k-1)}}] - \alpha P_{max} e^{\frac{\beta(\omega - \omega_p)}{\omega_p}} \tag{25.26}$$

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. It should be pointed out that this model, due to its simplicity, captures only the major aspect of an engine. Since much of the engine details are left out, the model exhibits only a moderate accuracy even with the correction term. We also recognize that the concept of this Bernoulli's principle-based model is not new and similar discussion can be found in existing work such as [35].

25.4 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 of the following four automotive engines: 2008 Mercedes CLS, 2006 Honda Civic, 2006 Pagani Zonda, and 1964 Chevrolet Corvair. Hopefully, these engines provide a good presence of vehicle makes, models, and model years. Listed in Table 25.1 are technical specifications of these engines. Additional information regarding parameter values used in this study is provided in Appendix B.

Table 25.1: Technical Spec of Engines Used in the Validation Study

Engine Tech Spec	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 (Nm)	531	174	750	209
ω at peak torque (rpm)	4000	4300	4050	2800
Engine volume (liter)	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

* values are estimated.

The primary criterion to evaluate these models is their accuracy. The following figures illustrate the relative performance of the three models using the empirical engine data as a benchmark. Each figure pertains to one of the engines and consists of two subplots - 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 difference from what they ought to be, so both power and torque curves are incorporated here for complete information.

In Figure 25.1, 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 over estimates 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 shifting to the left and translating upward.

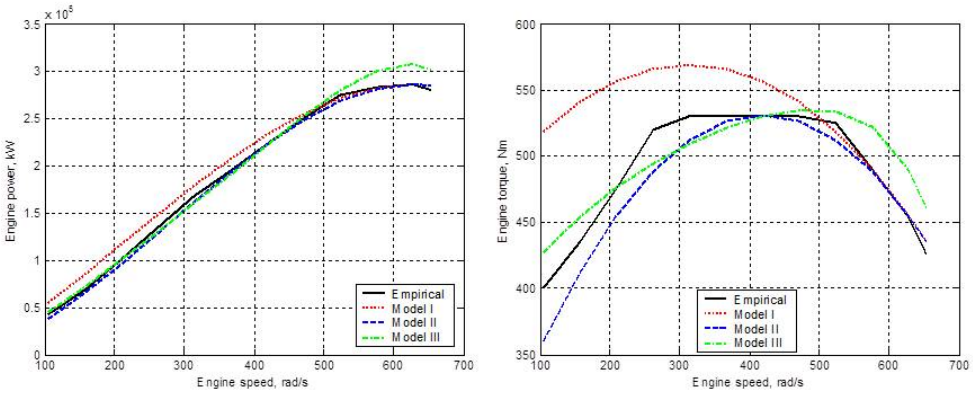


Figure 25.1: Model comparison based on 2008 Mercedes CLS engine

Figure 25.2 generally shows about the same pattern as that in Figure 25.1 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.

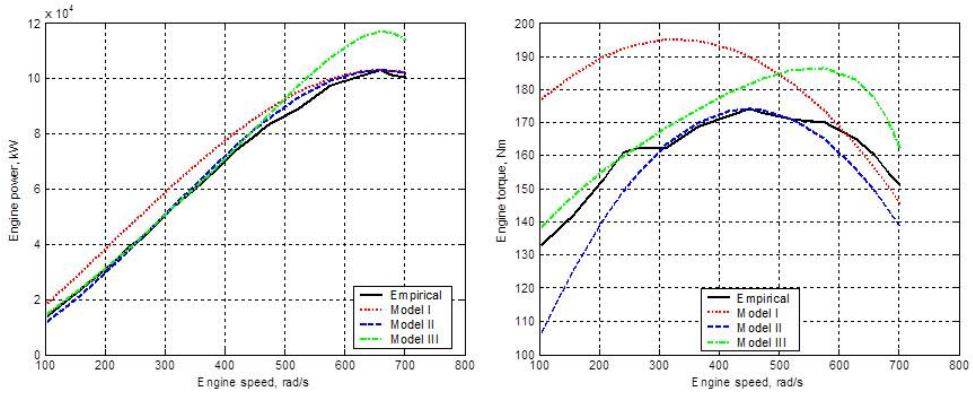


Figure 25.2: Model comparison based on 2006 Honda Civic engine

In Figure 25.3, 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.

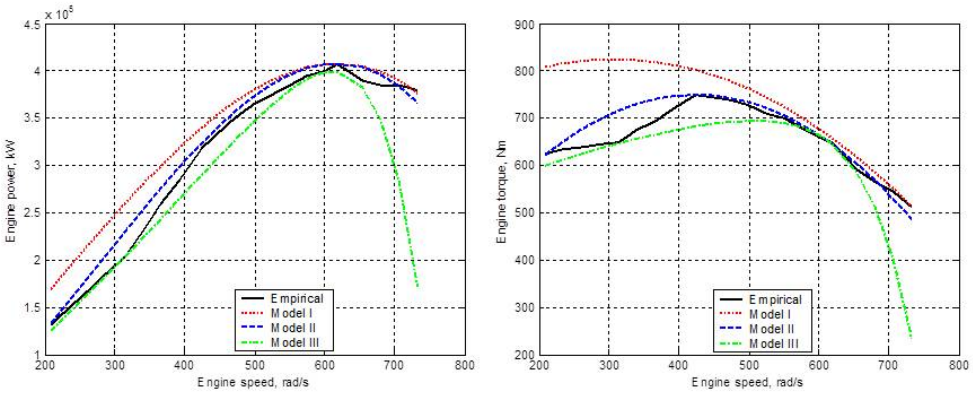


Figure 25.3: Model comparison based on 2006 Pagani Zonda engine

In Figure 25.4, Model II generally over estimates the torque before the peak torque. Except for a good fit of the peak torque, model I and III generally over estimate the torque.

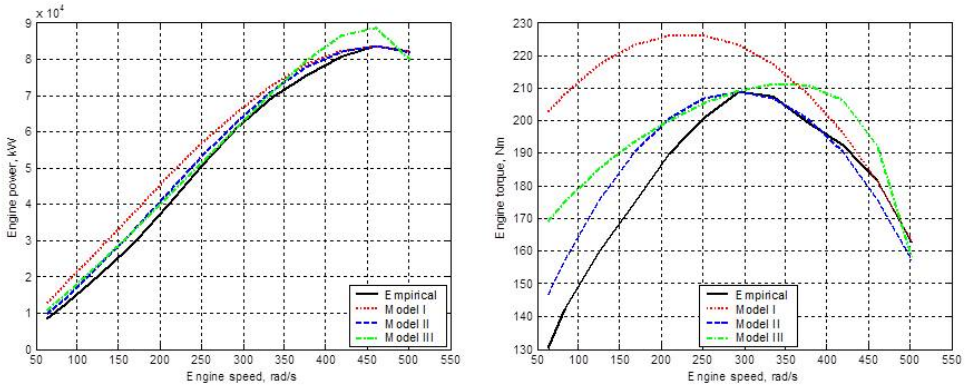


Figure 25.4: Model comparison based on 1964 Chevrolet Corvair engine

In order to quantify the accuracy of the three models, mean absolute percentage error (MAPE) is used as the measure of effectiveness. The MAPE is calculated as:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - X_i}{Y_i} \tag{25.27}$$

where n is number of samples, X_i is model estimate, and Y_i is the corresponding empirical value. Figure 25.5 confirms that Model II performs consistently well in both power and torque across the four engines. Its MAPE generally ranges between 3-7%. Though less than model II, Model III generally performs quite well, too, and its MAPE ranges between 4-9%. Model I performs the least in the three and its MAPE can be as high as 18%.

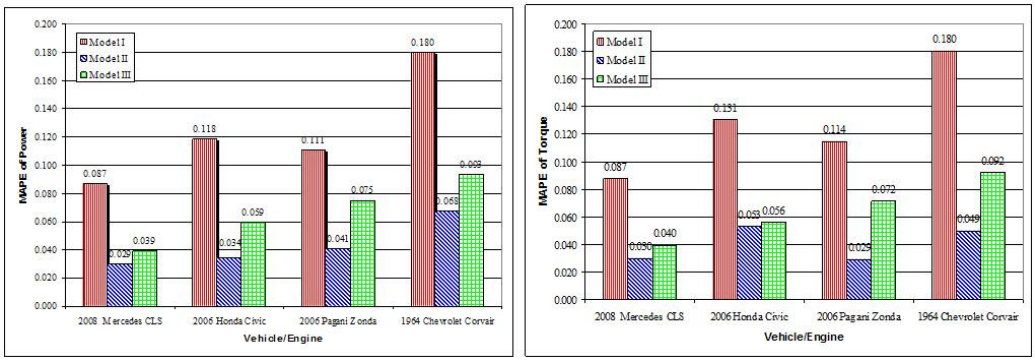


Figure 25.5: MAPE of the three models (left: power; right: torque)

The second criterion to evaluate these models is accessibility, i.e. the involvement of proprietary parameters and difficult-to-measure variables. In this regard, Model I and II are excellent because all they need are peak power and torque and their 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, model 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, their difference is pronounced in real-time applications, especially where a procedure has to be repeated very frequently. In terms of the fourth criterion - formulation, all the three models are analytical, no look-up table is involved. The fifth criterion is the need for calibration. In this regard, model I and II involve minimal calibration - all they need are peak power and torque and their associated engine speeds. Calibrating Model III is quite involved due to its proprietary parameter and calibration coefficients. The above comparison results are also highlighted in Appendix A. Overall, Model II outperforms the other two in terms of the above evaluation criteria.

25.5 Conclusion

An ideal engine model suitable for in-vehicle applications such as cooperative driving assistance system (CDAS) is expected to equip with reasonable accuracy, excellent computational efficiency, high accessibility, analytical formulation, and least calibration. Toward these goals, this paper presents three simple engine models: model I is an existing one and model II and III are developed by the authors. These models are formulated, validated, and evaluated. In terms of accuracy, model II and III are moderate while model I is low. In terms of computational efficiency, the three models are all acceptable with model I and II being particularly efficient. In terms of accessibility, model I and II are excellent because they do not require any proprietary parameter or difficult-to-measure variable. All the three models are equally good in terms of analytical formulation. Model III requires much effort on calibration, whole Model I and II involve minimal calibration. Overall, model II appears the best among the three in terms of all the evaluation criteria.

25.6 Appendix

25.6.1 A cross-comparison of engine models

25.6.2 Parameter Values

Engine efficiency: $\eta = 0.29$
Fuel energy density: $E_f = 46900000$ j/kg
Stoichiometric air-fuel ratio $\lambda = 0.068$
Air density $\rho = 1.29$ kg/m^3
Atmospheric pressure $p = 101325$ Pa
$\pi = 3.14159$
Heat capacity ratio of ideal gas $k = 1.407$
Molar mass of air $m' = 28.9$
Universal gas constant $R = 8314.5$ (Nm)/(kgmolK)
Coefficients in Model III: $\alpha = 0.15$ and $\beta = 10$

Table 25.2: A cross-comparison of engine models

Model	Accuracy	Complexity	Accessibility	Applications
I [44]	low	low	high	Vehicle dynamics
II	moderate	low	high	CVT-enabled in-vehicle control
III	moderate	moderate	moderate	Vehicle dynamics
[51]	low	low	moderate	Vehicle dynamics
[141]	moderate	low	low	Autonomous Cruise Control
[16]	moderate	moderate	low	Engine control analysis
[121] [27] [28]	moderate	moderate	low	Engine control analysis
[96] [161]	high	high	low	Engine control algorithms
[14]	high	high	N/A	Powertrain controllers and dynamics
[5]	high	high	low	Electric throttle control algorithm
[125]	high	high	low	Air-fuel ratio control and speed control
[135]	high	high	low	Engine diagnostics and control
[13]	high	high	low	The design procedure of internal combustion engines
[83]	high	Very high	low	Up-front design of engines for noise and vibration targets

Chapter 26

Vehicle Modeling

This chapter is a reproduction of the work in reference [58].

26.1 Introduction

Recent development of the Dedicated Short Range Communication (DSRC) standard, the Vehicular Ad Hoc Networks (VANET) technologies, and more broadly the Connected Vehicle Technology (CVT) have motivated many innovative applications such as Cooperative Collision Warning Systems and Cooperative Highway-Vehicle Automation. These applications are promising to address national transportation priorities such as traffic congestion which causes a waste of time and fuel totaling about $63.1 billion [1] and highway accidents which kill over 40,000 people [2] every year.

In order to facilitate the development of these applications and study their effects before large-scale deployment, a new generation traffic simulation tool is necessary for representing the new paradigm of traffic operations. For example, with CVT, vehicles will be able to communicate with other vehicles as well as with the transportation infrastructure. Thus drivers will receive more timely and accurate information. Such information helps drivers to take preventive actions and supports more compact platoons without sacrificing safety. In addition, inter-vehicular DSRC allows drivers to cooperate with each other to drive more safely and efficiently. However, conventional microscopic traffic simulation tools (such as CORSIM) are unable to represent enhanced information to drivers brought about by CVT and vehicle dynamic responses to improved driving strategies.

One of the dimensions to advance traffic simulation to the next generation is to incorporate vehicle dynamics so that the effects of improved driving strategies and the cooperation between drivers can be examined at a great level of detail. As an attempt toward this direction, this paper is devoted to the modeling of individual vehicle dynamics, utilizing a driver's desired acceleration, deceleration, and steering as inputs to determine vehicle dynamic responses including longitudinal acceleration, lateral acceleration, and yaw velocity. 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 CVT-enabled vehicles. Such an application involves large numbers of interacting and communicating vehicles, yet demands a modeling fidelity beyond the microscopic level. On this note, those vehicle models in automotive engineering are over-qualified given their complexity and high computation costs. Therefore, a Dynamic-Interactive-Vehicle (DIV) model with high computational efficiency and reasonable modeling fidelity is desirable.

The following section will briefly review existing efforts on representing vehicle dynamics in traffic simulation and this is followed by the development of the DIV model. After that, the DIV model is calibrated and validated. At the end, concluding remarks and a few future directions are presented.

26.2 Literature Review

Vehicular motion is primarily represented in traffic simulation via car-following models. There are many car-following models available today. Some of these models have a common structure which essentially has the response of the following vehicle as a function of a stimulus and a sensitivity factor. In general, the response refers to the acceleration of the following car as it tries to avoid colliding with the lead car while maintaining the driver's desired speed. The stimulus is often a function of the difference in distance and / or speed between the following and lead car. The sensitivity factor determines the weight of the response as a function of the following vehicle's speed and / or its distance behind the lead vehicle. Models with this structure that have been partly or fully implemented in traffic simulators including the Pipes

model [119] and the GHR / GM model [43]. What is noteworthy here is that these models do not directly account for the acceleration / deceleration capabilities of the vehicles they are representing. Instead maximum and minimum acceleration rates of the vehicles being represented are specified by users of these models.

As researchers attempted to improve the representation of car-following behavior, models presented in [46], [9] and [55] incorporated measures designed to capture realistic acceleration performance. The work presented in [128], [127], [129] highlights the formulation and comparison of a vehicle dynamics model which is capable of successfully predicting maximum acceleration performance. The model takes into account the effect of the vehicle's engine force and other dynamic properties along with aerodynamic, rolling, and grade resistances.

The realism with which vehicle motion is represented in traffic simulation has increased over the years through the use of existing car-following and simplified vehicle dynamics models. Further improvement may incorporate vehicle braking and steering to capture vehicle longitudinal and lateral movements, rather than only modeling the longitudinal movement in existing models. Additional improvement may also be made by modeling driver-vehicle interaction to examine the effect of enhanced information to drivers. Extending past research efforts in the modeling of vehicle dynamics will not only lead to improved understanding of traffic dynamics but also provide insights into highway design, traffic safety, more accurate estimates of vehicle emissions, and transportation forensics.

26.3 Development of the DIV Model

This section formulates the DIV model by considering the above-mentioned improvements.

26.3.1 Overview of the DIV Model

The DIV model will be capable of accepting three inputs from a vehicle driver: throttle position, brake pedal position, and the steering angle. The model will relate each input to a particular driver's desire and represent these various desires on a scale of 0 to 1 for the throttle and brake positions, and -1 to +1 for the steering angle. Each of these parameters will then interact with their

corresponding mechanisms to produce motion. The following subsections will present how DIV model treats the various components of a vehicle in order to faithfully capture its motion. These components include the engine, the braking system and the steering mechanism. Details of how the DIV model will account for effects due to rolling resistance, air resistance and gravity will also be presented in the following subsections.

26.3.2 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 using Newton second law of motion:

$$\sum F = m\ddot{x} = F_e - F_b - R_a - R_r - R_g \tag{26.1}$$

where:

W	=	weight of the vehicle (kg)
g	=	acceleration due to gravity (m/s^2)
F_e	=	tractive force produced by the engine (N)
F_b	=	force produce by the brake (N)
R_a	=	aerodynamic Resistance (N)
R_r	=	rolling resistance (N)
R_g	=	grade resistance (N)

Modeling Acceleration Performance

An engine plays an important role in vehicle acceleration performance. Here we adopt an engine model in [109], though it was found later that another model appeared to be better. The adopted model is briefly described below. The mass of air flowing into the engine's cylinders, m_a, is equal to:

$$m_a = \rho V_a \tag{26.2}$$

where:

ρ	=	density of the air in the combustion chamber (kg/m^3)
V_a	=	Total volume of the air in the combustion chamber of the engine (m^3)

However ρ and V_a need to be calculated first. The volume of air in all the combustion chambers of the engine, V_a, is essentially the product of the engine's displacement and its speed; resulting in:

$$V_a = V_e \cdot \frac{1}{2} \cdot \frac{\omega}{2\pi} \tag{26.3}$$

where:

V_e = engine displacement (m^3)

ω = engine speed (rad/s)

As for the density of air is the combustion chamber, ρ, it was calculated based on Bernoulli's Principle, which in essence states that an increase in the speed of a fluid results in a decrease in the pressure or gravitational energy experienced by that fluid as long as there is no work being done on the fluid. In calculating ρ, the cross-sectional areas of the inlet manifold and the opened throttle, along with their respective air pressure and densities were used. Therefore, after several iterations:

$$\rho = \rho_0 - \frac{\rho_0^2 V_a^2}{2A_0^2 p_0}(\frac{1}{\theta} + \frac{A_0^2}{A_2^2} - 2) \tag{26.4}$$

where:

A_0 = cross-sectional area of the inlet manifold (m^3)

A_2 = cross-sectional area of the inlet valve (m^3)

ρ_0 = density of air flow before the throttle (kg/ m^3)

p_0 = air pressure before the throttle (N/ m^2)

θ = throttle position (percent of throttle opening)

After computing the mass of air being taken into to the engine block the corresponding amount of fuel maybe estimated using the Stoichiometric air-fuel ratio (λ) for gasoline, which is 6.8%. Having a value for the amount of fuel entering into the engine, the amount of power generated by this quantity of fuel is determined by using the energy fuel density of gasoline, E_f, which is 46.9 MJ/kg. It is known that the efficiency of an internal combustion engine, η, is not 100% and is reflected in subsequent calculations. The effective power, P_{eff}, delivered to the vehicle's power-train mechanism is defined as:

$$P_{eff} = \eta \lambda E_f m_a \tag{26.5}$$

With the effective power calculated the effective torque, T_{eff}, delivered to the wheels of the vehicle can be determined with the following relationship:

$$T_{eff} = \frac{P_{eff}}{\omega} \tag{26.6}$$

Using the effective torque being delivered to the wheel, the effective engine force, F_e, produced by the engine to promote vehicle motion can therefore be calculated with the aid of the appropriate final transmission gear ratio, N_{ft}, and wheel radius, r.

$$F_e = \frac{T_{eff} N_{ft}}{r} \tag{26.7}$$

Modeling Braking Performance

The brake system will be 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 [94]. 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 [45]. Therefore, using the results from this study, the DIV model will be able to not only account 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 maximum deceleration rate through minimal pedal force. Using this proportionality constant, the following formulation will be used in the DIV model to represent the brake system of a vehicle and the driver's interaction with that system.

$$F_b = d_b \cdot p_f \cdot W \tag{26.8}$$

where:

F_b = Brake Force (N)
d_b = driver's desire to brake (0-1)
p_f = pedal-force gain coefficient

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:

$$R_a = \frac{\rho}{2} C_D A_f v_r^2 \tag{26.9}$$

where:

ρ = Mass density of the air (0.07651 lb/ft^3- performance test condition)

C_D = Coefficient of aerodynamic 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 by:

$$F_g = \pm W sin\theta \approx W\theta \tag{26.10}$$

where: θ = Grade of the incline in radians.

26.3.3 Modeling Lateral Movement

The structure used to represent the movement of the DIV model in the X-Y plane was adapted from [93], which included the formulation a kinematics and a dynamic framework to model a vehicle's motion in a two-dimensional space. The kinematics framework that was presented in the aforementioned manuscript was chosen for the DIV model for two primary reasons: 1) all the pertinent dynamic properties of the vehicle were already accounted by other means in the DIV model, and 2) the ease of use with accurate X-Y position representation.

At the base of the kinematics framework for the 2-D representation of vehicle motion is the treatment of the vehicle as a non-homonymic system, which is a system that does not guarantee return to its original position, even if its original configuration is reached. Along with the non-homonymic treatment of the vehicle, non-homonymic constraints, which are related to the velocity of the vehicle, are held 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 of low speeds is negligible. The general form of the non-holonomic constraint maybe represented as:

$$\dot{x}sin(\phi) - \dot{y}cos(\phi) = 0 \tag{26.11}$$

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 26.1 for an illustration of the coordinate

system being used and also the definition of variables that will be used in the development of the DIV model.

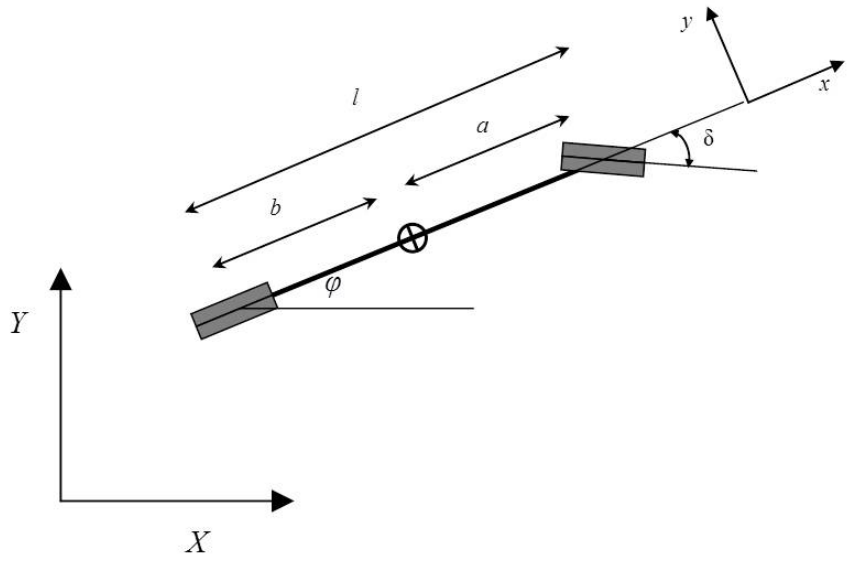


Figure 26.1: DIV model in the X-Y plane

After a few more iterations of Equation 26.11 the velocity of the center of gravity with respect to the global coordinate system is defined as:

$$\dot{X} = \dot{x}cos(\phi) - \dot{y}sin(\phi) \tag{26.12}$$

$$\dot{Y} = \dot{x}sin(\phi) + \dot{y}cos(\phi) \tag{26.13}$$

Having Equations 26.12 and 26.13, 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 change the orientation of the vehicle.

$$\dot{y} = \dot{\phi}b \tag{26.14}$$

$$\dot{\phi} = fractan(\delta)l\dot{x} \tag{26.15}$$

And

$$\delta = d_t \frac{\pi N_s}{r_s} \tag{26.16}$$

where:

d_t	=	Driver's desire to turn (-1 to +1)
N_s	=	# of steering wheel revolutions
r_s	=	Steer ratio (ratio of radians dialed to Ackerman angle)

26.4 DIV Model Calibration

A key feature of the DIV model is that it is meant to be easily calibrated. The calibration process of the DIV model will entail the user providing the model with 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 vehicles, for example Cars.com. The vehicle performance specifications that the DIV model requires include: 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 gradient of the roadway. Once the values of the vehicle performance specification and the various values describing the surrounding environment are entered into the DIV model, it will be able to replicate the motion of the vehicle.

26.5 DIV Model Validation

In validating 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 kph (60 mph) is recorded, as well as the time it takes a vehicle to cover a distance of 402 m (mile). The Federal Motor Carrier Safety Administration (FMCSA) dictates maximum allowable stopping distances from various speeds that all vehicle manufacturers must satisfy, standardizing vehicle braking. And finally, to measure how well a vehicle handles, the diameter of the circle scribed by the

vehicle's outer front wheel is recorded, after the maximum steering angle has been dialed.

26.5.1 Test Vehicles for Validating the DIV Model

To increase the applicability of the DIV model, passenger cars were chosen, representing approximately 58% of the registered passenger vehicles in the United States. Therefore, upon successful validation of the DIV model with the use of passenger cars, the DIV model will be capable of representing the majority of vehicles of today's roadways. Three different types of passenger cars were used in this validation process - a sports car - 2006 Porsche Cayman S, a large passenger car - 2006 Ford Fusion Sedan SE, and a small passenger car - 2006 Honda Civic Coupe EX.

26.5.2 Validating Acceleration Performance

As previously mentioned, the tests chosen to verify how well the DIV model replicates the acceleration performance of an automobile are the 0 - 97 kph (0 - 60 mph) test and the 402 m (mile) test. In these tests, the driver of the vehicle opens the throttle body to its maximum position and the time it takes the vehicle to get from rest to 97 kph (60 mph), as well as the time taken for the vehicle to cover a distance of a 402 m (mile) are recorded. The DIV model simulated these tests and its results were compared to the published results for the test vehicles of similar tests. Published results for the various tests were obtained from (Cars.com 2007). The comparisons of test results are presented in Figure 26.2.

Tests	**Porsche Cayman S**			**Ford Fusion**			**Honda Civic**		
	Obs.	*DIV*	*Error*	*Obs.*	*DIV*	*Error*	*Obs.*	*DIV*	*Error*
	(sec)	*(sec)*	*(%)*	*(sec)*	*(sec)*	*(%)*	*(sec)*	*(sec)*	*(%)*
0-97 kph (0-60 mph)	5.15	4.98	3.30	6.89	6.78	1.60	7.84	7.74	1.28
402 m (¼ mile)	13.67	13.20	3.44	15.47	15.05	2.71	16.08	15.89	1.18

Figure 26.2: Comparison of acceleration performance

Evaluation of Test Results

The absolute percentage error between the observed results and that of the DIV model provides a means of quantitatively validating the DIV model. For the purpose of this paper, an absolute percentage no greater than 5% between the observed results and that of the DIV model represents a successful attempt by the DIV model in replicating the acceleration performance of a real vehicle. As seen from the above figure, the absolute percentage error ranges from 1.60% to 3.44%, signifying that fact that the DIV model is successful at replicating an automobile's acceleration performance.

To further highlight the validity of the DIV model, as it attempts to represent the acceleration performance of an automobile, a diagonal plot was created. The diagonal plot provides a means of qualitatively evaluating how well the DIV model replicated acceleration performances. In essence, for the diagonal plot, if the simulated mechanism, in this case the DIV model, replicated the observed results then the ideal fit will be a 45^o line. As seen from Figure 26.3, this is almost the case, once again proving the validity for the DIV model when replicating acceleration performance.

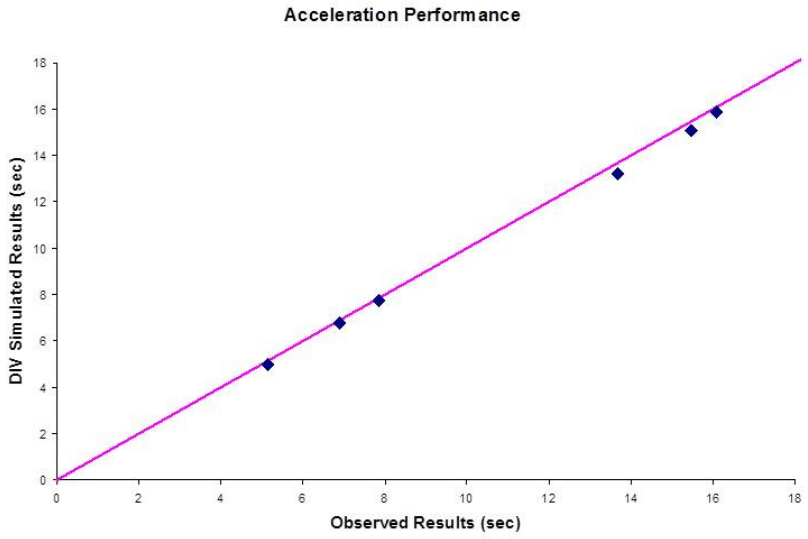


Figure 26.3: Diagonal plot of observed vs. simulated results

26.5.3 Validating Deceleration Performance

Due to the complexity of the braking system and the various ways of measuring braking performance the DIV model's braking capabilities will be validated by comparing its performance to the safety standards set by the Federal Motor Carrier Safety Administration. The braking performance of the DIV model will be compared to Part 571 of the Federal Motor Vehicle Safety Standards - Standard No. 105, which describes the requirements for Hydraulic and Electric Brake Systems [36].

There are four effectiveness tests geared toward ensuring safe braking performance.

- Effectiveness Test 1: a small vehicle / passenger car, with pre-burnished brakes, should be able to come to rest from 48 kph (30 mph) and 97 kph (60 mph) within 17 m (57 ft) and 66 m (216 ft) respectively.
- Effectiveness Test 2: a passenger car should be able to rest from speeds of 48 kph (30 mph), 97 kph (60 mph) and 129 kph (80 mph) within 16 m (54 ft), 62 m (204 ft) and 117m (383 ft) respectively.
- Effectiveness Test 3: a lightly loaded passenger car should be able to come to rest from 97 kph (60 mph) within 59 m (194 feet).
- Effectiveness Test 4: a passenger car should be able to come to rest from speeds of 48 kph (30 mph), 97 kph (60 mph), 129 kph (80 mph) and 161 kph (100 mph) within distances of 17 m (57 ft), 66 m (216 ft), 123 (405 ft), and 205 m (673 ft) respectively.

Note that when referring to brakes being pre-burnished and burnished, the DIV model will not account for the difference between the two. This largely due to the complexity in treating the two stages of how worn the brakes are and the lack added fidelity that would be gained upon including the effects of pre-burnished and burnished brakes. Another point of clarification is the definition of a lightly loaded vehicle. According the Federal Motor Carrier Safety Administration a lightly loaded vehicle is the unloaded vehicle weight plus 181 kg (400 lbs).

Evaluation of Tests Results

In carrying out the various effectiveness tests on the DIV model, after the speed, specified by the test, is reached the driver applies maximum brake

force. The stopping distance is then calculated based on the distance travel between the times the driver applies the brake to when the car comes to rest. In testing the braking performance of the DIV model, the four effectiveness tests were conducted on the DIV model's representation of the three test vehicles and the results summarized in Figure 26.4.

It is important to clarify the meaning of the numbers in Standard No. 105 in order to evaluate the success of a model in terms of its ability to replicate the braking performance of real vehicles. Note that these numbers are *upper bounds* (i.e. the maximum allowable distances) set by the Federal Motor Carrier Safety Administration to ensure that car makers produce braking systems that can bring cars to rest from a series of initial conditions within a specified distance after the driver has applied the brake. Therefore, a model is successful if it produces a stopping distance that is less than or no greater than 5% of the stopping distance specified for a particular effectiveness test.

In Figure 26.4, ✓ and **x** are used to indicate whether or not the DIV model was successful at representing the braking performance of a particular vehicle during a specific effectiveness test. The ✓ represents a successful replication of the braking performance and the **x** represents a failure. The above evaluation of the test results was a quantitative means of validating the model. For a qualitative evaluation of the model Figure 26.5 illustrates a diagonal plot of the Standard No. 105 upper bounds versus the braking performance results from the DIV model.

As discussed above, points above the 45^o line in this diagonal plot represent instances where the DIV model failed to produce stopping distances that are within the standards set by the FMCSA, while points on or below the line represents instances of success.

Examining both the quantitative and qualitative evaluations of the test results of the DIV model's ability to replicate braking performance, one will notice that in light of the several instances where the DIV failed to produce stopping distances under the aforementioned boundaries, the model still performs relatively well. This level of success in the midst of the instances of failure provides motivation for additional research, to correct the problems that may lie in the DIV model.

26.5.4 Validating Lateral Movement

To illustrate this validation process the 2006 Porsche Cayman S lateral movement test is presented. The turning circle as defined by Road and Track Mag-

Braking Performance	Standard No. 105	DIV	Error	Success
	Stop Dist. m (ft)	Stop Dist. (ft)	%	
Porsche Cayman S				
Effectiveness Test #1	17 (57)	17.12 (56.18)	-1.4	✓
	66 (216)	60.97 (200.04)	-7.4	✓
Effectiveness Test #2	16 (54)	17.12 (56.18)	4.0	✓
	62 (204)	60.97 (200.04)	-1.9	✓
	117 (383)	99.06 (325)	-15.1	✓
Effectiveness Test #3	59 (194)	60.92 (199.87)	3.0	✓
Effectiveness Test #4	17 (57)	17.12 (56.18)	-1.4	✓
	66 (216)	60.97 (200.04)	-7.4	✓
	123 (405)	99.06 (325)	-19.8	✓
	205 (673)	143.17 (469.72)	-30.2	✓
Ford Fusion				
Effectiveness Test #1	17 (57)	18.79 (61.67)	8.2	✗
	66 (216)	64.87 (212.83)	-1.5	✓
Effectiveness Test #2	16 (54)	18.79 (61.67)	14.2	✗
	62 (204)	64.87 (212.83)	4.3	✓
	117 (383)	103.70 (340.23)	-11.2	✓
Effectiveness Test #3	59 (194)	64.91 (212.97)	9.8	✗
Effectiveness Test #4	17 (57)	18.79 (61.67)	8.2	✗
	66 (216)	64.87 (212.83)	-1.5	✓
	123 (405)	103.70 (340.23)	-16.0	✓
	205 (673)	147.33 (483.35)	-28.2	✓
Honda Civic				
Effectiveness Test #1	17 (57)	68.06 (68.06)	19.4	✗
	66 (216)	72.37 (237.44)	9.9	✗
Effectiveness Test #2	16 (54)	68.06 (68.06)	26.0	✗
	62 (204)	72.37 (237.44)	16.4	✗
	117 (383)	116.59 (382.50)	-0.1	✓
Effectiveness Test #3	59 (194)	71.25 (233.77)	20.5	✗
Effectiveness Test #4	17 (57)	68.06 (68.06)	19.4	✗
	66 (216)	72.37 (237.44)	9.9	✗
	123 (405)	116.59 (382.50)	-5.6	✓
	205 (673)	156.89 (514.74)	-23.5	✓

Figure 26.4: Braking performance results and analysis

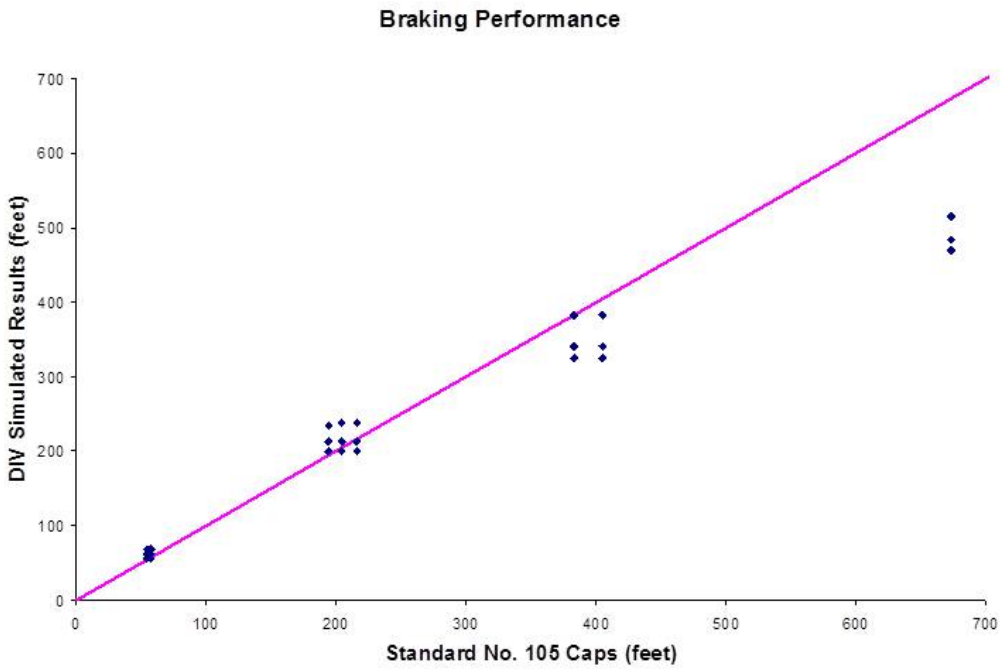


Figure 26.5: Diagonal plot of observed vs. simulated braking distances

azine is the circle that is scribed by the outside front tire when the maximum steering lock dialed in. The measurement that is often used to represent the turning circle of a given automobile is the diameter of that circle. According to the manufacture's specifications, the 2006 Porsche Cayman S, with a steering ratio of 15.5:1, scribes a turning circle with an approximate diameter of 11.1 m (36.4 ft).

Using all the specifications of the Porsche Cayman S in the DIV model and dialing the maximum steering lock, the diameter of the turning circle is approximately 11.23 m (36.87 ft). The diameter determined by the DIV model has to be converted from the diameter of the circle scribed by the center of gravity 8.58 m (28.15 ft.) to the diameter of the circle scribed by the outer front wheel of the vehicle. To do this, the distance from the center of gravity to the center of the front axle is calculated, doubled 2.66 m (8.72 ft.) and then added to the diameter of the circle scribed by the center of gravity.

Similar procedures were conducted to determine how well the DIV model is capable of reproducing the turning circles of the other two test vehicles. The DIV model's results of the diameters of the vehicles' turning circles and the corresponding absolute percentage errors are presented in TFigure 26.6.

Turning Circle	**Porsche Cayman S**			**Ford Fusion**			**Honda Civic**		
	Obs.	*DIV*	*Error*	*Obs.*	*DIV*	*Error*	*Obs.*	*DIV*	*Error*
	(m)	*(m)*	*(%)*	*(m)*	*(m)*	*(%)*	*(m)*	*(m)*	*(%)*
Diameter	11.09	11.24	1.29	12.19	12.86	5.47	10.79	10.29	4.60

Figure 26.6: Diameter of turning circle - DIV model vs. real vehicles

In terms of validating this aspect of the DIV model, a successful representation of the lateral movement of an automobile is had when the absolute percentage error in the DIV model's output of the diameter of a turning circle is no greater 5% of the published corresponding value. Given the percentage error in the above figure, it can be deduced that the DIV model demonstrate a rather high level of success when replicating a vehicle's lateral movement.

26.6 Conclusion and Future Directions

The DIV model is a simple and computationally affordable vehicle dynamics model which is able to represent vehicle motion with a reasonable fidelity. Upon validating the DIV model, its capabilities of replicating three sets of performance test results for three passenger cars were demonstrated. The three performance tests measure how well a vehicle accelerates, brakes, and turns.

The acceleration test included a measure of how long it takes a vehicle to accelerate from rest to 97 kph (60 mph) and also how long it takes a vehicle to cover 402 m (mile). The DIV model output times whose absolute percentage errors are no greater than 3.5% of the observed values of the three test vehicles - well within the 5% range criteria used to represent successful replication of acceleration performance. As for the brake performance test, the DIV model outputs the required distances to bring a vehicle to rest from particular speeds. These values of stopping distances were compared to a series of standards (upper bounds) set by the Federal Motor Carrier Safety Administration (FMCSA). In several instances the DIV model was able to output distances within the limits set by the FMCSA; however, there are few instances in which the DIV model's output exceeded the standards. The Porsche Cayman representation was the most successful, while the Honda Civic representation was the least successful. In the performance test to determine how well the DIV model represents lateral movement, we calculated the diameter of the circle scribed by the outer front wheel of the vehicle after the maximum steering angle was dialed. The absolute percentage error between the DIV model's results and those measured from real passenger cars were no more than 5.5%.

The DIV model realistically replicates the two-dimensional movement of an automobile, while allowing interactions with both the driver and the road. In addition, due to its simple mathematical representation, the DIV model is very computationally efficient involving little calibration efforts.

Future directions of this research effort can be focused on the following specific areas: the incorporation of the preferred engine model in reference [109], the mathematical representation of a vehicle's braking system, the comprehensive treatment of the lateral movement of a vehicle, and the validation procedure in determining how well the vehicle model is capable of representing vehicle dynamics.

The Braking System: Currently, the DIV model treats the brake system of

a vehicle according to the result from a Driver-Vehicle Braking Performance study conducted in 1970 (Mortimer et al. 1970). Not only is this study due for an update given the advances in the brake technology used on today's vehicles, but also this treatment of the braking system is not vehicle specific. The next step for the DIV model in representing the brake system of a vehicle is to utilize simplified mathematical representations of how a driver places her foot on the brake pedal and produce a force on the wheels, retarding the motion of a vehicle.

Lateral Vehicle Movement: The DIV does a relatively good job of representing the lateral movement of a vehicle at low speeds. When representing a vehicle's lateral motion the DIV model does not account for slippage at the tire-road interface. At low speeds slippage is negligible and does not affect the motion of a vehicle. But at a high speeds, tremendous slippage can occur and greatly influences the motion of a vehicle. Therefore the next step for the DIV model is to account for slippage at the tire-road interface.

Validation Procedure: The current performance tests involved in the validation procedure only test performances with maximum driver input, i.e. maximum gas and brake pedal displacement and maximum steering angle. The next step would be to conduct performance test on both a real vehicle and the DIV model where maximum driver input is rare.

Chapter 27

The Field Theory

27.1 Motivation

Research on highway traffic flow over the past half a century has resulted in many follow-the-leader theories, each of which was proposed with its own motivation. For examples, in the GM family of models [11, 43], 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; Pipes model [119], Forbes model [39, 38, 37], and Gipps model [46] were inspired by safe driving rules; in Psycho-physical models [89, 157], driver reactions were triggered by perception thresholds; Rule-based models [73, 76] was motivated by the fuzzy logic in driver decision making. Though the motivation behind some other car-following models such as [99] and equilibrium traffic flow models such as [50, 49, 146, 29] might not be clear, they were so formulated due to their reasonable performances. Two questions naturally arise. First, would it be possible to have a unifying framework that coherently interpret and relate these models? Second, would it be possible to root such a unifying framework in first principles so that traffic flow theory is furnished with a solid foundation and connected to other branches of sciences and engineering?

Chapters 27, 28, 29, and 30 are motivated by the above questions. Chapter 27 attempts to address the second question, Chapter 28 is intended for the first question, Chapter 29 provides some supporting evidences, and Chapter 30 presents a multi-scale modeling perspective. In this chapter, our attention shall be devoted to the modeling of driver operational control in a transportation system, i.e. the motion and interaction of driver-vehicle units

on a long homogeneous highway. Based on first principles (e.g. physical laws and social rules), a phenomenology is postulated which represents the driving environment perceived by a subject driver as a 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 to seek the least hazardous route by navigating through the field along its valley.

27.2 Physical Basis of Traffic Flow

Three systems are of particular interest: a physical system, a transportation system, and a social system, illustrated in Figure 27.1. The physical system typically consists of non-living 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 vary 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 goes the transportation system which involves both living entities (human drivers) and non-living 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. In addition, it is close to the soft end when strategic planning is concerned, while it migrates toward the hard end if tactical decision 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. Below are some examples of such similarities:

27.2.1 Mechanics Phenomena

In Physics, forces are the cause of change of motion. In addition, they are measurable and their effects reproducible. For example, Newton's second law of motion stipulates that the velocity of an object changes if it is subject to a non-zero external force; Newton's third law says that, for every action, there

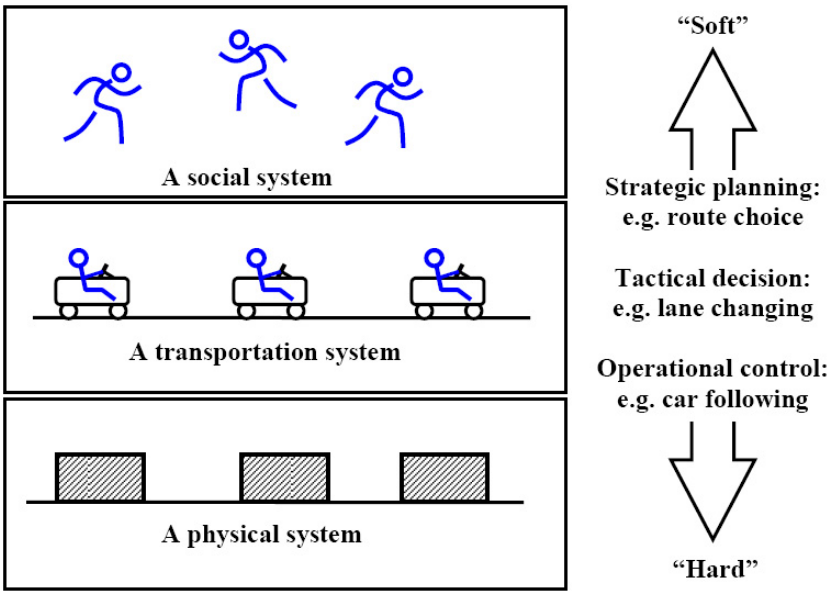


Figure 27.1: Three systems

is an equal and opposite reaction. Similarly, "forces" exist in traffic flow, but such forces are subjective matters. Consequently, they are non-measurable 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, he or she speeds up or gives his or her 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 pre-determined direction much like free objects always fall to the ground. The reason why free objects fall is because they are constantly subject to earth 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 mentally in 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 vs. streets).

M2: Free flow An object in free fall will accelerate to an equilibrium speed due to air resistance, 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 oppose 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 condition, the choice may vary over different drivers or within 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 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 later case resides in the driver 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 his or her 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 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 would perceive some stress which motivates him or her to steer back in 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 would 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, 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 an illustration of the barriers in Figure 27.2. It should be noted that these barriers are not real objects, but only imaginary in drivers' minds.

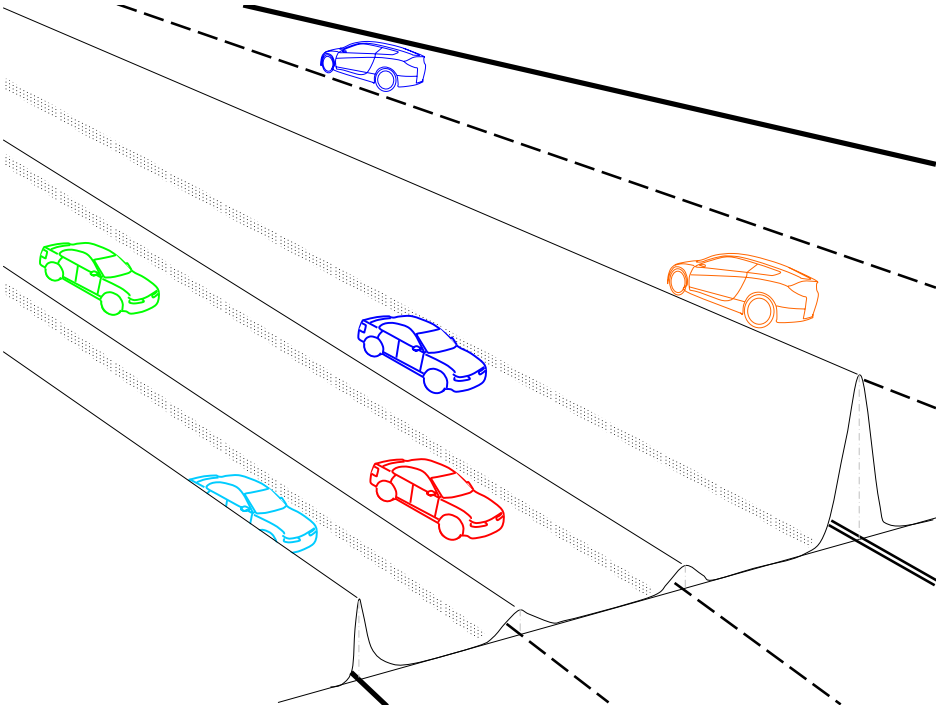


Figure 27.2: Road barriers

27.2.2 Electromagnetic Phenomena

An object can exert a force on another object in either of the following two ways: collision and action at a distance. For example, hitting a ball with a bat is the former and finding a needle using a magnet is the latter. Though collisions are not uncommon on highways, action at a distance is what vehicles normally interact with each other and examples of this kind include some of the above mechanics phenomena as well as the following:

E1: car following Back to the example in the beginning of Section 27.2 where a fast vehicle catches up with a slow vehicle, the fast driver perceives

an imminent collision if he or she keeps driving at that 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 respond, 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 distance 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 Continuing the above example and assuming that the fast vehicle tailgates (i.e. following at a dangerously short distance), it is likely that the opposite force is perceived by the slow driver who may respond by speeding up or giving his or her way to the fast follower. Back to the analogy, 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 takes effect on the person. In addition, a driver's ability to perceive depends on where he or she scans and how frequent 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.

27.2.3 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 km per hour), and density (in vehicles per km), while the wave is determined by frequency (in Hz 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.

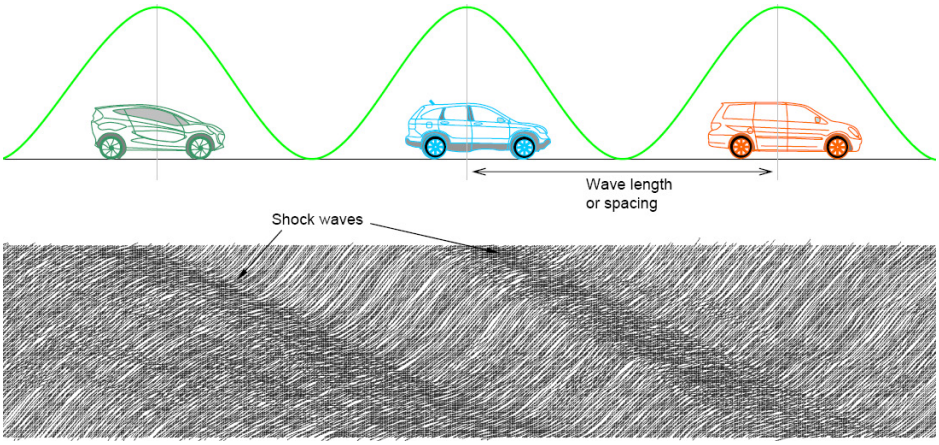


Figure 27.3: Traffic and waves

The upper part of Figure 27.3 shows a platoon of vehicles as a harmonic wave.

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 catch 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 27.3 illustrates a few shock waves observed in vehicle trajectories.

W3: Wave-particle duality All matter, particularly small-scale objects, exhibits both wave-like 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 exhibit wave property (e.g. kinematic waves and shock waves).

27.2.4 Statistical Mechanics Phenomena

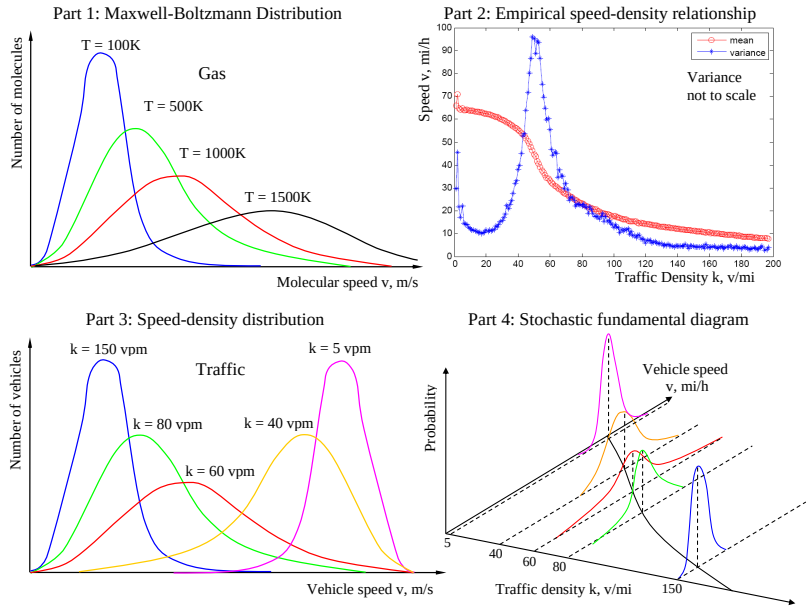


Figure 27.4: Maxwell-Boltzmann distribution in traffic flow

Traffic flow has been modeled by many as a one-dimensional compressible fluid, such as a gas. In gases, the speeds of gas molecules follow Maxwell-Boltzmann distribution, see an illustration in Part 1 of Figure 27.4. Remarkable in the distribution is the increase in average speed and speed variance as temperature becomes higher. In contrast, traffic flow exhibits a different trend. Empirical observations in Part 2 of the figure show that the variance of traffic speed peaks around optimal density (where capacity flow occurs) and drops at both ends. Such a distribution is illustrated in Part 3 of the figure in contrast to Maxwell-Boltzmann distribution and further elaborated in a three-dimensional model in Part 4 which forms the basis of a stochastic fundamental diagram.

27.3 The Field Theory

Given the above similarities between the transportation system and the physical system, a phenomenology of traffic flow (i.e. the Field Theory) is mo-

tivated 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 non-lining objects (e.g. roadways and vehicles), it is subject to both physical laws and social rules. As such, the phenomenology is formulated progressively based on a set of postulates, two of which (Postulates 1 and 3) are physical and another two (Postulates 2 and 4) social.

Postulate 1: A road is a physical field

Postulate 1 is motivated by Phenomena M1, M2, and M4 in Section 27.2. In the longitudinal (x) direction, a driver-vehicle unit is subject to a gravity along the road:

$$G_i = m_i \times g_i \tag{27.1}$$

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, i.e. $g_i = g_i(\Theta, \Lambda, \Xi)$.

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 desired speed v_i which in term is related to the free-flow speed of the road v_f (subscript f here means free flow, not a vehicle ID), i.e. $R_i = R_i(\dot{x}_i, v_i, v_f)$. Therefore the net force acting on unit i in the longitudinal direction can be expressed as:

$$m_i \ddot{x}_i = G_i - R_i \tag{27.2}$$

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 varnishes, 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 of Section 27.2, 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 his or her lane, see illustrations in Figures 27.2 and 30.4. The effect of such a force is to push the unit back to the center of the current lane. Based on 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:

$$N_i = -\frac{\partial U_i^R}{\partial y_i} \tag{27.3}$$

Postulate 2: A driver responds to his or her surrounding 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 non-contact forces which is 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 up. 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 front, it is the driver behind who is responsible for watching safety distances and held liable for a rear-end collision should it happen. Therefore, 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 surrounding vary with his or her viewing angle and scanning frequency, see Figure 27.5. For example, the area right 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

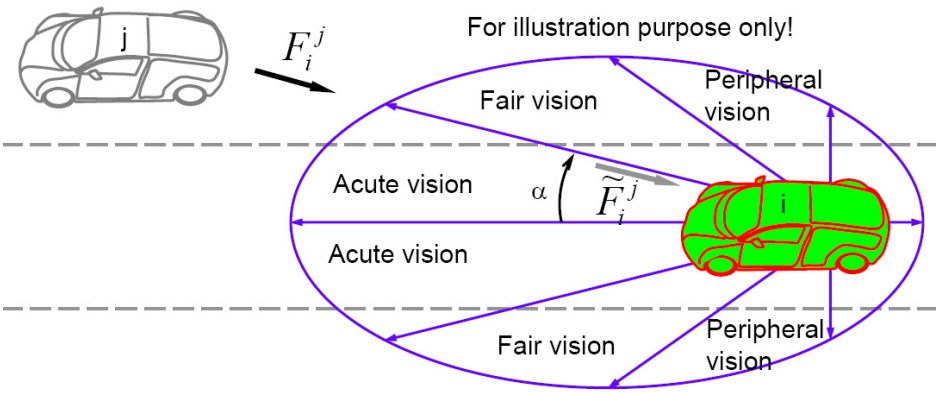


Figure 27.5: Distribution of driver's attention

driver's fair vision zones, the frontal areas in side lanes receive 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 these area when making a lane change. 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 make a lane change 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 the responsibility rests on drivers behind. Therefore, it is reasonable to assume that the driver's directional response to his or her surrounding, γ_i, is a function of his or her viewing angle α_i, i.e.$\gamma_i = \gamma(\alpha_i)$. 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, i.e.,

$$\tilde{F}_i = F_i \times \gamma(\alpha_i) \tag{27.4}$$

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., $\alpha_i = \pi$), respectively.

Postulate 3: A driver interacts with others by action at a distance

As described in E1, E2, and E3 in Section 27.2, a driver is able to sense the presence of other vehicles and obstacles in its 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 hazard of 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 30.4 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 repelling force it receives. The longitudinal section of the field is illustrated as the curve above the x axis.

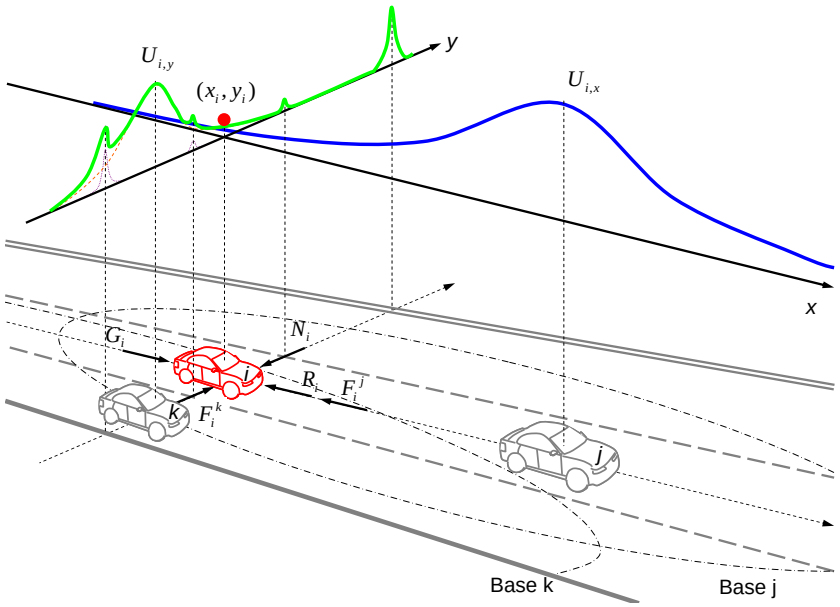


Figure 27.6: The illustration of a field

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 at the side lane of unit i, the influence is not in longitudinal x direction but in lateral y direction, i.e., 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, i.e. $U = U(\Theta, \Gamma)$ 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 the field it creates.

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 using 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, i.e.,

$$U_i = U_i^B + U_i^R + U_i^C \tag{27.5}$$

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, i.e. the least stressful route. For example, Figure 30.4 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:

$$F_i^j = -\frac{\partial U_{i,x}}{\partial x} \tag{27.6}$$

The effect of F_i^j is to push unit i back to keep safe distance. By incorporating the driver's unsatisfied desire for mobility $(G_i - R_i)$, the net force in the x direction can be determined as:

$$m_i \ddot{x}_i = \sum F_{i,x} = G_i - R_i - F_i^j = (m_i g_i - R_i) + \frac{\partial U_{i,x}}{\partial x} \tag{27.7}$$

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:

$$m_i \ddot{y}_i = \sum F_{i,y} = F_i^k - N_i = -\frac{\partial U_{i,y}}{\partial y} \tag{27.8}$$

By incorporating time t, driver i's perception-reaction time τ_i, and driver i's directional response γ_i, Equations 27.7 and 27.8 can be expressed as:

$$m_i \ddot{x}_i(t+\tau_i) = \sum \tilde{F}_{i,x}(t) = \gamma_i^0[G_i(t) - R_i(t)] + \gamma(\alpha_i^j)\frac{\partial U_{i,x}}{\partial x} \tag{27.9a}$$

$$m_i \ddot{y}_i(t+\tau_i) = \sum \tilde{F}_{i,y}(t) = -\gamma(\alpha_i^k)\frac{\partial U_{i,y}}{\partial y} \tag{27.9b}$$

where $\gamma_i^0 \in [0,1]$ represents the unit's attention to its unsatisfied desire for mobility (typically $\gamma_i^0 = 1$), α_i^j and α_i^k are viewing angles which are also functions of time. The above system of equations summarizes the phenomenology in generic terms and constitutes the basic law governing a unit's motion on a planar surface.

27.4 A Special Case

Though the generic form of the phenomenology is able to explain some traffic phenomena qualitatively, rigorous modeling of traffic flow requires a specific form which is the focus of this section. 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, Occam's razor turns out to be a good rule of thumb which 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, to which empirical observations are a good test (this is responded in Chapter 29), (2) it should take a simple functional form (responded in Equation 29.1) that involves physically meaningful parameters but not calibration coefficients (responded in Section 27.5), and (3) it should provide a sound microscopic basis to aggregated behavior, i.e. macroscopic equilibrium models (responded in Section 27.5 and Chapter 29). With these considerations in mind, some simplifications are made to the generic theory as the first step to formulation a specific form:

Simplification 1 Rather than formulating the field itself, the specific form formulates forces resulted from the field directly.

Simplification 2 The specific form decouples equations in the longitudinal x and lateral y directions, i.e. the longitudinal equation is used to model driver's longitudinal control (e.g. car-following behavior) and the lateral equation is only used when a lane change is to be considered.

Simplification 3 Directional response $\gamma(\alpha)$ is treated as follows: for car following, the subject driver only responds 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.

27.4.1 Motion in the Longitudinal Direction

With these simplifications, vehicle motion in the longitudinal direction is formulated as follows. Again, time t and response delay (i.e. perception-reaction time τ) are dropped for convenience.

Unsatisfied desire for mobility

The term $(G_i - R_i)$ explains a driver-vehicle unit's unsatisfied desire for mobility. Intuitively, when a unit starts from stand still, i.e., $\dot{x}_i = 0$, its unsatisfied desire for mobility is the greatest. As the unit speeds up, $(G_i - R_i)$ decreases accordingly but is still positive, i.e., it still accelerates the unit to higher speeds. When the unit achieves its desired speed, i.e., $\dot{x}_i = v_i$, 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 over-satisfied 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:

$$G_i - R_i = m_i g_i [1 - (\frac{\dot{x}_i}{v_i})^{\delta}] \tag{27.10}$$

where δ is a calibration parameter.

Action at a distance

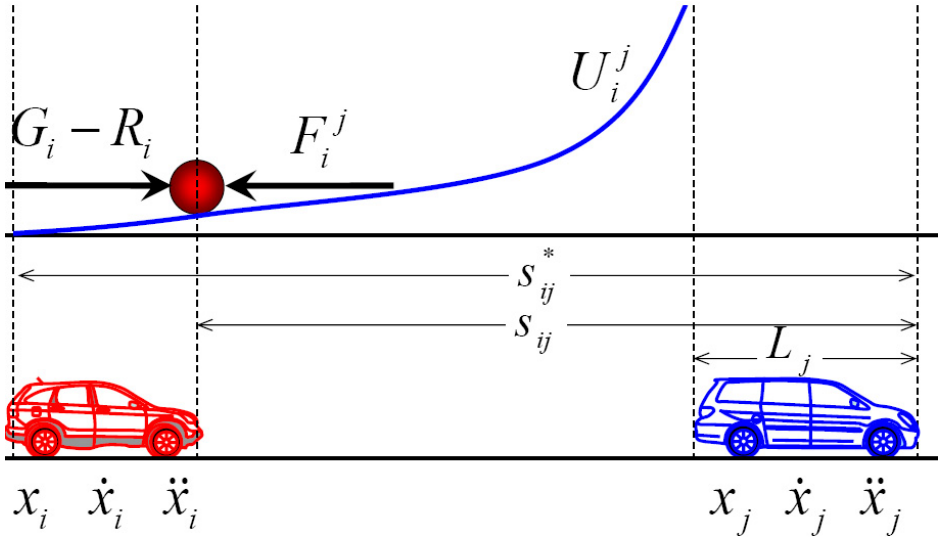


Figure 27.7: Action at a distance

When a fast unit i (with displacement x_i, speed $\dot{x}_i$, and acceleration $\ddot{x}_i$) catches up a slow unit j (with x_j, $\dot{x}_j$, and $\ddot{x}_j$), the former is subject to a non-contact force, F_i^j, from the latter. Such a non-contact 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, it takes effect when unit i becomes close (e.g. within the range of its desired spacing s_{ij}^*), it increases as the spacing becomes even shorter ($s_{ij} \downarrow$), and it goes to 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 base line, 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 can be:

$$F_i^j = f(e^{s_{ij}^* - s_{ij}}) \tag{27.11}$$

where $s_{ij}^* - s_{ij}$ represents how far unit i intrudes into s_{ij}^*.

The Longitudinal Control Model

Combining the above, the effort that is required by driver i to control his or her vehicle in the longitudinal direction can be expressed as:

$$m_i\ddot{x}_i = G_i - R_i - F_i^j = m_i g_i[1 - (\frac{\dot{x}_i}{v_i})^\delta] - f(e^{s_{ij}^* - s_{ij}}) \tag{27.12}$$

or

$$m_i\ddot{x}_i = m_i g_i[1 - (\frac{\dot{x}_i}{v_i})^\delta - f(e^{s_{ij}^* - s_{ij}})] \tag{27.13}$$

if the coefficient of term F_i^j is chosen properly. Though Equation (27.13) can be further detailed in many possible ways, the following special case is of particular interest. Putting time t and response delay τ back on and eliminating vehicle mass m at both sides, Equation (27.13) can take the following special form:

$$\ddot{x}_i(t + \tau_i) = g_i[1 - (\frac{\dot{x}_i(t)}{v_i})^\delta - e^{\frac{s_{ij}(t)^* - s_{ij}(t)}{Z}}] \tag{27.14}$$

If one chooses $\delta = 1$ and $Z = s_{ij}(t)^*$, the above equation reduces to a more specific form:

$$\ddot{x}_i(t + \tau_i) = g_i[1 - (\frac{\dot{x}_i(t)}{v_i}) - e^{\frac{s_{ij}(t)^* - s_{ij}(t)}{s_{ij}(t)^*}}] \tag{27.15}$$

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 Section 27.5), (2) it makes physical and empirical sense (see Chapter 29), and (3) it provides a sound microscopic basis to aggregated behavior, i.e. the corresponding equilibrium model (see Chapters 28 and 29).

The determination of desired spacing $s_{ij}(t)^*$ admits safety rules. An example safety rule can be that one should leave a time gap of at least one perception-reaction time in front to ensure safety. As such, the desired spacing can be formulated as

$$s_{ij}^*(t) = \dot{x}_i(t)\tau_i + L_j \tag{27.16}$$

Another example safety rule can be the following: the desired spacing should allow vehicle i to stop behind its leading vehicle j after a perception-reaction time τ_i and a deceleration process (at rate $b_i > 0$) should vehicle j applies an emergency brake (at rate $B_j > 0$). After some math, the desired spacing can be determined as follows:

$$s_{ij}^*(t) = \frac{\dot{x}_i^2(t)}{2b_i} + \dot{x}_i\tau_i - \frac{\dot{x}_j^2(t)}{2B_j} + L_j \tag{27.17}$$

Alternatively, the desired spacing can be determined by admitting other safety rules which are not enumerated here. It is interesting to note that empirical results suggest that the Longitudinal Control Model performs better if perception-reaction time τ is replaced by a speed-varying one $\tau' = \tau e^{-\dot{x}_i/v_i}$, which might be justified by the effect that drivers tend to be more vigilant when traveling at higher speeds.

27.4.2 Motion in the Lateral Direction

A driver-vehicle unit's motion in the lateral y direction involves decision at two 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 27.8 illustrates the scenario of Figure 27.7 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 the figure positions unit 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 road barrier (lane line). Driven by its desire for mobility, unit i climbs up onto U_i^j, during which 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-road. Hopefully, an opportunity exists in the left lane because the elevation of unit i (where the ball rides) is higher than both 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 onto the front of U_i^p, the effect of which is shown in the middle part of Figure 27.8.

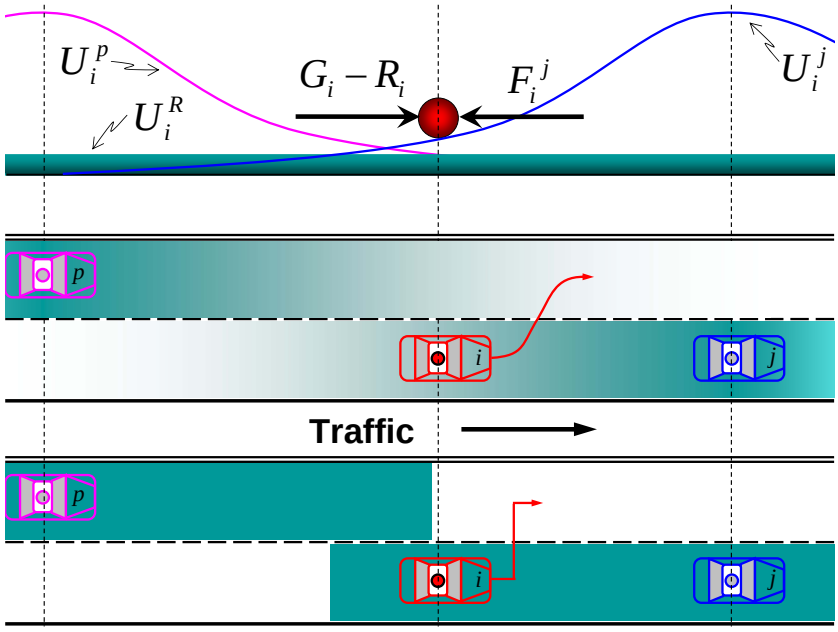


Figure 27.8: Lane change

The above Lateral Control Model can be further simplified by reducing a smooth field to a flat "personal space" into which intruding by another unit is undesirable. For example, the bottom part of Figure 27.8 illustrates units j and p's personal spaces after eliminating 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 which 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.

27.5 Properties of the Theory

27.5.1 Model Parameters

Parameters involved the Longitudinal Control Model (Equation 29.1) are: roadway gravity g, perception-reaction time τ, desired speed v, minimum safety room L, deceleration b, and emergency deceleration B. More specifically:

Roadway gravity Equation (27.10) reveals that the roadway gravity g_i perceived by driver i is actually the maximum acceleration that the driver is willing to apply when starting from stand still. Meanwhile, this parameter is constrained by vehicle physical properties and its value is publicly available in automobile magazines such as Motor Trend.

Desired speed The desired speed varies among different drivers and on different roads. Even for the same driver, his or her desired speed may vary over time and under different situations. Therefore, the desired speed of driver i can be modeled as a stochastic process $\{v_i(t)|t \geq 0\}$ which is typically slow-varying over time t. During a particular period of time, $v_i(t)$ may be treated as time invariant and, hence, be represented as a sample from an underlying distribution $\{v_i \in \Phi_v(\mu_v, \sigma_v^2)\}$ where μ_v is mean and σ_v is standard deviation. In addition, μ_v can be related to the free-flow speed, v_f, of a particular road, e.g. $\mu_v = v_f$.

Perception-reaction time Again, this parameter typically varies among drivers and, for the same driver, varies over time and under different situations. Data of this parameter may come from human factors studies using driving simulators and its value is typically around one to two seconds.

Minimum safety room This is the minimum safe distance between two vehicles and can be treated as the nominal vehicle length (typically 5-10 m or 16-32 ft), i.e. the actual vehicle length plus a little buffer space at each end.

Deceleration rates The traffic engineering profession typically uses a comfortable deceleration of 3.4 m/s^2. The emergency deceleration depends

on a vehicle's physical property and is typically around $6 - 8\ m/s^2$.

More discussion on model parameters are presented in Chapter 29.

In general, model parameters can be roughly grouped in three categories: (1) parameters with physical meaning and easy to calibrate, such as perception-reaction time, (2) parameters with physical meaning but difficult to calibrate such as optimal density (i.e. traffic density when capacity flow occurs), and (3) calibration coefficients without physical meaning. Obviously, Equation (29.1) only involves parameters in the first category, which is desirable in mathematical modeling.

27.5.2 Tentative definition of two vague terms

With the phenomenology and its special case, 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, i.e.

$$M_i(t) = \frac{\dot{x}_i(t)}{v_i} \tag{27.18}$$

where $\dot{x}_i(t)$ is driver i's actual speed and v_i 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:

$$M_i = \frac{1}{T_i} \int_0^{T_i} \frac{\dot{x}_i(t)}{v_i} dt \tag{27.19}$$

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:

$$M = \frac{1}{N}\sum_{i=1}^{N}(\frac{1}{T}\int_0^{T_i}\frac{\dot{x}_i(t)}{v_i}dt) \tag{27.20}$$

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, i.e. $C = 1 - M$ which is expressed relatively in 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, i.e.

$$C_i(t) = U_i(t) \tag{27.21}$$

As defined in Equation 27.5, 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, one would experience no congestion if he or she moves on a roadway that is free of impedance due to other moving units, need for lane changes, and traffic control devices. Therefore, any increase of these would add to one'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:

$$C_i = \int_0^{T_i} C_i(t)dt \tag{27.22}$$

Further, the congestion experienced by all drivers in a traffic system can be calculated as the sum of personal congestion over the driver population:

$$C = \sum_{i=1}^{N} C_i \tag{27.23}$$

27.5.3 Multiple Regimes

In addition to describing car-following behavior, the Longitudinal Control Model is able to represent multiple regimes encountered by a driver moving on the road:

- Start-up: vehicle i starts moving from stand still.
- Speed-up: after start-up, vehicle i continues to accelerate to higher speeds.
- Free flow: as vehicle i speeds up, it settles at its desired speed if it is unimpeded.
- Cutting off: a sudden decrease in inter-vehicle spacing due to another vehicle j squeezing in in front.
- Following: vehicle i has to adopt j's speed so as to follow the leader.
- Stop-and-go: vehicle i is forced to stop and go due to j's brief stopping.

- Trailing: vehicle i is following a speeding leader.
- Approaching: vehicle i is getting close to a slower or stationary leader.

- Stopping: vehicle i tries to stop behind a stationary object separated by a minimum spacing.

More discussion on these regimes and a demonstration of the model's capability to handle them is presented in the later chapter.

27.5.4 Steady-State Behavior under Equilibrium

If a system is in steady state, any property of the system is unchanging in time. More specifically, a traffic system in steady state would consists of homogeneous vehicles which exhibit unform behavior over time and space. Therefore, under steady-state condition, vehicles lose their identities (e.g. $\tau_i \rightarrow \tau$), vehicles travel at uniform speed (i.e. $\dot{x_i} = \dot{x_j} = v$ and $\ddot{x_i} = 0$), and drivers' desired speeds converge to the free-flow speed of the road (i.e. $v_i \rightarrow v_f$). As such, the special case in Equation (29.1) can be aggregated to the following equilibrium equation:

$$v = v_f[1 - e^{1-\frac{k^*}{k}}] \tag{27.24}$$

where $s_{ij} \rightarrow s = \frac{1}{k}$, $s^*_{ij} \rightarrow s^* = \frac{1}{k^*}$, $s^* = \frac{v^2}{2b} + v\tau - \frac{v^2}{2B} + L$, b average deceleration, B average emergency deceleration, $L = \frac{1}{k_j}$ average minimum spacing, and k_j jam density. If one ignores the difference in braking (i.e. $b = B$) and considers the effect that drivers are more vigilant at higher speeds, a more specific form of Equation (28.11) is resulted:

$$v = v_f[1 - e^{1 - \frac{1}{k(v\tau' + L)}}] \tag{27.25}$$

The above equation describes an equilibrium speed-density relationship at the macroscopic level. Unfortunately, the equation has speed v at both sides, which is undesirable because this makes it difficulty to express speed v as an explicit function of density k, as conventional equilibrium speed-density relationships do. However, this is unavoidable if the model were to make physical sense and the reason is the following. The essence of the model is to use desired spacing s^*_{ij} as the base line, see Figure 27.7. If the trailing vehicle i is outside of s^*_{ij}, it is virtually not influenced by its leading vehicle j. When vehicle i intrudes into s^*_{ij}, the driver is under stress to observe safety rule and such a stress is represented as a force which is exponentially related to the amount of intrusion (i.e. $s^*_{ij} - s_{ij}$), see Equation (29.1). Note that the desired spacing is speed-varying because driver's safety need is directly related to their speeds, as stipulated in Equation (29.2). Therefore, the speed term on the right-hand side of Equation (29.3) is inherited from its microscopic counterpart. With this understanding, it becomes clear that an attempt to eliminate speed dependence on the right-hand side in order to achieve mathematical attractiveness inevitably has to trade off the model's empirical accuracy and microscopic basis.

While an explicit speed-density function is unavailable, an explicit density-speed relation does exist. Solving Equation (29.3) for k yields:

$$k = \frac{1}{(v\tau' + L)[1 - ln(1 - \frac{v}{v_f})]} \tag{27.26}$$

Figure 27.9 illustrates the fundamental diagram resulted from the above equilibrium model. Four plots are generated: speed-density (v-k), speed-flow (v-q), flow-density (q-k), and speed-spacing (v-s). Remarkable in this figure are: boundary conditions, capacity condition, and general shape of each plot.

Boundary conditions Two extreme points in the v-k plot are $(k = 0, v = v_f)$ and $(k = k_j = 1/L, v = 0)$ which can be verified in Equations (29.3) and

(29.4). The slope of the v-s plot can be determined by taking derivative of v with respect to $s = \frac{1}{k}$ in Equation (29.4):

$$\frac{dv}{ds} = \frac{1}{\tau e^{-v/v_f}(1 - \frac{v}{v_f})[1 - ln(1 - \frac{v}{v_f})] + \frac{v\tau e^{-v/v_f} + L}{v_f - v}}$$

The slope of the v-s curve is found by evaluating $\frac{dv}{ds}$ at point $(s = L, v = 0)$:

$$\frac{dv}{ds}|_{(s=L,v=0)} = \frac{1}{\tau + \frac{L}{v_f}}$$

Capacity condition Capacity flow (i.e. $q = q_m$ achieved when $k = k_m$ and $v = v_m$) can be determined by setting derivative of $q = kv$ with respect to either k or v to zero. Unfortunately, the solution is difficult to find analytically, but hopefully solvable numerically.

Shapes of curves The plots in Figure 27.9 resemble closely to those curves typically observed in empirical data (see empirical results in Chapter 29). In particular, both q-k and v-q curves are concave, which are desirable.

27.5.5 Connection to Existing Knowledge Base

The purpose of this subsection is to place the phenomenology in a broader context and show how the theory relates to existing knowledge base.

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. The next chapter is specifically devoted to this topic.

Connection to Other Engineering Disciplines

Lennard-Jones potential plays an important role in engineering, particularly in granular flow and molecular dynamics. In molecular dynamics, computer

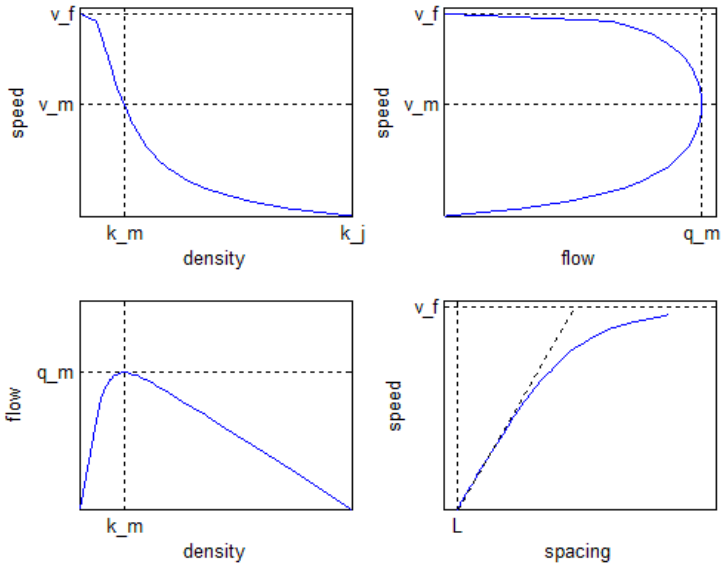


Figure 27.9: The fundamental diagram

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. Lennard-Jones potential is the underlying model to determine the motion and interaction of these particles. In material engineering and granular flow, 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 27.10, the Lennard-Jones potential takes the following form:

$$U(r) = 4\epsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}] \tag{27.27}$$

where U is 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 inter-particle 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 (27.12), takes a similar form.

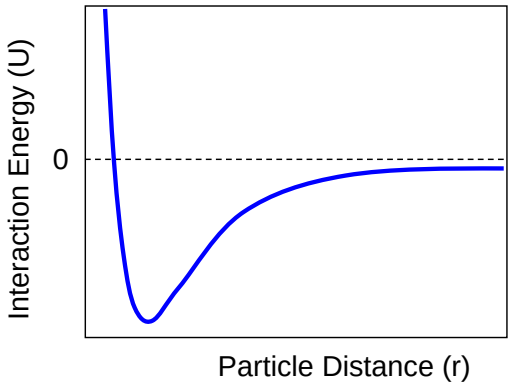


Figure 27.10: Lennard-Jones potential

For example, the long-range attraction ($m_i\ddot{x}_i = m_i g_i[1 - (\frac{\dot{x}_i}{v_i})]$) is due to 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 the two 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 is originated from empirical observations and, in return, able to explain real-world phenomena. Rather than being derived directly from first principles, the phenomenology is formulated in a way that is consist with fundamental theory. For example, the essence of 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. For another example, Equation (27.9) is a special form of Newton's second law of motion if one ignores 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 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 hard collision, gives rise to 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 mean free path (i.e. desired car-following distance s_{ij}^*) and 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.

27.6 Conclusion

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. As a matter of fact, the transportation system does exhibit many physical properties, which partly motivates the proposed phenomenology of traffic flow.

To pave the foundation of the phenomenology, some physical phenomena in traffic flow are analyzed in relation to its social properties, in particular motivations of drivers' decision observed from their driving experiences. These phenomena, including those of mechanics, electromagnetic, wave, 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 phenomenology is progressively formulated based on a series of postulates, two of which are physical and another two social. The first postulate (physical) assumes that a roadway is a physical field in which a vehicle is subject to a roadway gravity and meanwhile a resistance due to driver's willingness to observe traffic rules (e.g. speed limits). The second postulate (social) accounts for driver's directional responsiveness to his or her surrounding. 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 hazard of collision. The distinction between a perceived field 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, i.e. a driver always tries to achieve gains (e.g. mobility and safety) and avoid losses (e.g. collisions and violation of traffic rules). Combining the above postulates, the phenomenology is generically formulated as a system of equations governing the motion of a vehicle on a roadway in relation to other vehicles.

To facilitate the application of the phenomenology, a special case is derived from the generic theory based on some simplifications. The resulting special case consists of a Longitudinal Control Model with specific functional form and a Lateral Control Model governing lane changing behavior. Properties of the special case are discussed including its model parameters, multiple driving regimes it is capable of describing, its steady-state behavior at the macroscopic level, and its connection to existing knowledge base.

Part V

The Unified Perspective

Chapter 28

The Unified Diagram

28.1 Motivation

Half a century ago, Newell [99] proposed a nonlinear car-following model of the following form:

$$\dot{x}_i(t+\tau_i) = v_i(1 - e^{-\frac{\lambda_i}{v_i}(s_{ij}(t)-L_i)}) \tag{28.1}$$

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 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 Newell model and furnish it with a possible motivation (this section). In doing so, it is found that the interpretation gives rise to a broader picture that can be used to relate some existing traffic flow models to each other (Section 28.2). As such, a unified perspective can be casted on traffic flow modeling with bridges not only within but also between microscopic and macroscopic levels (Section 28.3).

Without further delay, Newell model can be slightly rearranged as follows:

$$1 - \frac{\dot{x}_i(t+\tau_i)}{v_i} - e^{\frac{L_i - s_{ij}(t)}{v_i/\lambda_i}} = 0 \tag{28.2}$$

The above equation in turn is a special case of the following equation when vehicle i's acceleration $\ddot{x}_i$ is zero:

$$\ddot{x}_i(t+\tau_i) = g_i[1 - \frac{\dot{x}_i(t)}{v_i} - e^{\frac{L_i - s_{ij}(t)}{v_i/\lambda_i}}] \tag{28.3}$$

where g_i is a positive, non-zero parameter associated with vehicle i. Equation (28.3) is a *dynamic* car-following model which describes the acceleration performance of vehicle i, whereas Equation (28.2) is a *steady-state* version of the dynamic model since the former describes the speed choice of driver i in steady state, i.e. 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/hr (or 70 mi/hr)"), such a goal is achieved over time, during which 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 is naturally resulted from the driver's operation on gas and brake. 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 and remains there to bring his or her vehicle to a stop, a deceleration is constantly applied till the vehicle stops, at which moment the deceleration suddenly disappears. In contrast, the speed profile of the vehicle has to be smooth, i.e. starting from its initial speed and continuously decreasing to zero. For another example, when a subject vehicle is being cut off from its leader (due to 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, even though the limiting acceleration may be involved, it is still within the scope of dynamic modeling.

Equation (28.3) can be further arranged as follows by multiplying both sides by vehicle mass m_i:

$$m_i\ddot{x}_i(t+\tau_i) = m_i g_i - m_i g_i \frac{\dot{x}_i(t)}{v_i} - m_i g_i e^{\frac{L_i - s_{ij}(t)}{v_i/\lambda_i}} \tag{28.4}$$

One immediately recognizes that the above equation takes the form of Newton's second law of motion:

$$\sum F_i = G_i - R_i - F_i^j \tag{28.5}$$

where $\sum F_i = m_i \ddot{x}_i(t + \tau_i)$, $G_i = m_i g_i$, $R_i = G_i \frac{\dot{x}_i(t)}{v_i}$, $F_i^j = G_i e^{\frac{L_i - s_{ij}(t)}{v_i/\lambda_i}}$. Therefore, Equation (28.4) 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 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 non-contact force, it is an action at a distance as if it were mediated by a "field".

28.2 A Broader Perspective Based on Field Theory

Extending the above discussion, it appears that it is appropriate to interpret driver's operational control using the concept of 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 modeling of vehicle motion is to seek the least hazardous route by navigating through the field along its valley. The Field Theory of such a nature has been proposed in Chapter 27. Only major results of the Field Theory are reproduced below for easy reference.

28.2.1 Highlight of the Field Theory

The generic form of the Field Theory is:

$$\begin{cases} m_i\ddot{x}_i(t+\tau_i) & = \gamma_i^0[G_i(t) - R_i(t)] + \gamma(\alpha_i^j)\frac{\partial U_{i,x}}{\partial x} \\ m_i\ddot{y}_i(t+\tau_i) & = -\gamma(\alpha_i^k)\frac{\partial U_{i,y}}{\partial y} \end{cases} \tag{28.6}$$

Readers are referred to Chapter 27 for derivation and notation of the results 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:

$$\ddot{x}_i(t+\tau_i) = g_i[1 - (\frac{\dot{x}_i(t)}{v_i})^\delta - e^{\frac{s_{ij}(t)^* - s_{ij}(t)}{Z}}] \tag{28.7}$$

or

$$\ddot{x}_i(t+\tau_i) = g_i[1 - (\frac{\dot{x}_i(t)}{v_i}) - e^{\frac{s_{ij}(t)^* - s_{ij}(t)}{s_{ij}(t)^*}}] \tag{28.8}$$

if one chooses $\delta = 1$ and $Z = s_{ij}(t)^*$. $s_{ij}(t)^*$ is motivated by safety rules and can take many forms, of which two examples are provided:

$$s_{ij}^*(t) = \dot{x}_i(t)\tau_i + L_j \tag{28.9}$$

$$s_{ij}^*(t) = \frac{\dot{x}_i^2(t)}{2b_i} + \dot{x}_i(t)\tau_i - \frac{\dot{x}_j^2(t)}{2B_j} + L_j \tag{28.10}$$

Aggregating Equation 29.1 over vehicles by assuming steady-state conditions yields the following equilibrium speed-density relation:

$$v = v_f[1 - e^{1-\frac{k^*}{k}}] \tag{28.11}$$

A more specific form can be:

$$v = v_f[1 - e^{1-\frac{1}{k(v\tau' + L)}}] \tag{28.12}$$

or

$$k = \frac{1}{(v\tau' + L)[1 - ln(1 - \frac{v}{v_f})]} \tag{28.13}$$

28.2.2 Relating microscopic car-following models

Following the rationale in Section 28.1, it turns out that the Field Theory could be used as a framework to relate traffic flow models and such an effort could extend to models at both microscopic and macroscopic levels.

Newell Model

Back to Newell nonlinear car-following model. Comparing Equations 29.8 and 28.7 reveals that the former can be resulted if one chooses to: (1) apply steady-state condition, i.e. $\ddot{x}_i(t+\tau_i) = 0$, (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 29.8 is the tangent of the speed-spacing curve (Figure 1 in Newell's original paper) evaluated when speed is zero. This parameter can be interpreted as the reciprocal of 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 (See Part I Subsection 5.3). With this understanding, the vehicle interaction force F_i^j suggested by 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 i at his or her desired speed v_i during one perception-reaction time τ_i.

Continuing the above effort, the Field Theory is related to other microscopic car-following models as follows.

Forbes Model

Forbes model [39, 38, 37] 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 29.1) as a means to determine the *desired spacing* $s^*_{ij}(t)$ which is formulated in Equation 28.9.

General Motors (GM) Models

The family of GM models [11, 43] is generically formulated in its fifth model (GM5):

$$\ddot{x}_i(t+\tau_i) = \alpha \frac{\dot{x}_i^m(t+\tau_i)[\dot{x}_j(t) - \dot{x}_i(t)]}{[x_j(t) - x_i(t)]^l} \tag{28.14}$$

If one chooses $m = l = 1$, Equation 28.14 reduces to the fourth generation GM model (GM4):

$$\ddot{x}_i(t+\tau_i) = \alpha \frac{\dot{x}_i(t+\tau_i)[\dot{x}_j(t) - \dot{x}_i(t)]}{[x_j(t) - x_i(t)]} \tag{28.15}$$

where x_i, $\dot{x}_i$, $\ddot{x}_i$, τ_i are the displacement, speed, acceleration, and perception-reaction time of the subject vehicle i respectively; similar notations apply to its leader j; α is a dimensionless coefficient. In relation to the Field Theory, GM4 model considers only the vehicle interaction force F_i^j and ignores unsatisfied desire for mobility $(G_i - R_i)$, see Chapter 27 for details. Rather than translating intrusion exponentially to vehicle interaction force as in the Longitudinal Control Model (Equation 29.1), 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 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 29.5, Vehicle i is attracted to (or repelled by) vehicle j if the latter travels faster (or slower) than the former.

Gipps Model

The [46] model consists of a system of two inequalities with one governing free-flow regime and the other car-following.

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.

$$\dot{x}_i(t+\tau_i) = \dot{x}_i(t) + 2.5 g_i \tau_i (1 - \frac{\dot{x}_i(t)}{v_i}) \sqrt{0.025 + \frac{\dot{x}_i(t)}{v_i}} \tag{28.16}$$

The above equation can be re-written in the following differential form after considering time difference τ:

$$\ddot{x}_i(t+\tau_i) \approx \frac{\dot{x}_i(t+\tau_i) - \dot{x}_i(t)}{\tau_i} = g_i'(1 - \frac{\dot{x}_i(t)}{v_i}) \tag{28.17}$$

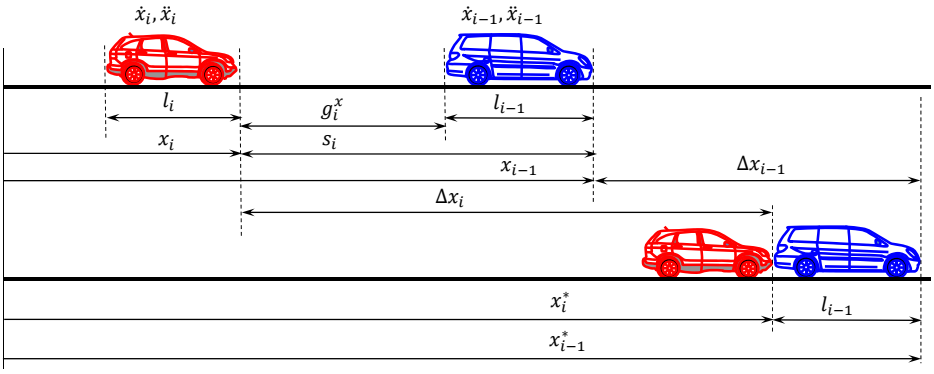


Figure 28.1: The Gipps model

where $g_i' = 2.5 g_i \sqrt{0.025 + \frac{\dot{x}_i(t)}{v_i}}$. Note that Equation 28.17 is actually the *unsatisfied desire for mobility* term in Equation 29.1 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 its 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 applies an emergency brake $(B_j > 0)$. The scenario is illustrated in Figure 28.1 and the model reproduced below:

$$s_{ij}^*(t) \geq \frac{\dot{x}_i(t+\tau_i)^2}{2b_i} + \frac{\tau_i}{2}[\dot{x}_i(t) + \dot{x}_i(t+\tau_i)] + \dot{x}_i(t+\tau_i)\theta - \frac{\dot{x}_j^2}{2B_j} + L \quad (28.18)$$

The astute reader has recognized that the above model describes the *desired spacing* which follows exactly the same safety rule used to derive Equation 29.2 which is slightly modified from and simpler than the above model. Of course, one could opt to use this model in place of Equation 29.2 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 have to be resulted 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) [145, 55] is expressed as a superposition of the follower i's acceleration term and a deceleration term which depends on the desired spacing s_{ij}^*:

$$\ddot{x}_i(t+\tau_i) = g_i[1 - (\frac{\dot{x}_i}{v_i})^\delta - (\frac{s_{ij}^*}{s_{ij}})^2] \tag{28.19}$$

where δ is acceleration exponent, $s_{ij} = x_j - x_i$ is the spacing between vehicle i and its leader j, and 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 29.1, IDM strikingly resembles the Longitudinal Control Model. From the perspective of the Field Theory, IDM relates the interaction F_i^j between vehicle i and its leader j to the squared ratio of desired spacing s_{ij}^* to actual spacing s_{ij}. In addition, 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 [147, 148] combines the [119] and [50] models into a single equation:

$$s_{ij} = c_1 + c_3\dot{x}_i + c_2/(v_f - \dot{x}_i) \tag{28.20}$$

where $c_1 = v_f(2v_m - v_f)/(k_j v_m^2)$, $c_2 = v_f(v_f - v_m)^2/(k_j v_m^2)$, $c_3 = 1/q_m - v_f/(k_j v_m^2)$, v_f is the free-flow speed of the roadway facility, k_j is the jam density, and v_m is the optimal speed occurred 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 was developed by [9] which consists of a set of acceleration algorithms (reproduced below applying our notation):

A1: Vehicle i is moving but has not yet reached its desired speed v_i. Depending on i's initial speed and urgency of the task, the acceleration rate is found by entering Tables 1 and 2 provided in the original paper.

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.

A3: Vehicle i was stopped and has to start from stand still. A maximum acceleration rate is applied constrained by a non-collision constraint after a response delay.

A4: Vehicle i is in car-following with its leader j. A_4 is determined by satisfying the following safety rule: vehicle i should leave a non-negative gap ($s_{ij} - L_j \geq 0$) from j should 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 other variables are as defined before.

A5: Vehicle i in car-following is subject to a non-collision constraint which is reinforced by considering the desired spacing: $s^*_{ij}(t) = x_j(t) - x_i(t + \Delta t) \geq max\{\dot{x}_i(t+\Delta t)\tau_i + L_j$ 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 $\dot{x}_i(t+\Delta t) = \dot{x}_i(t) + A_5 \Delta t$, B_i and B_j are 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 Forbes model [39, 38, 37] and the second choice is similar to that of [46] model if driver i is willing to apply emergency brake (i.e. $b_i = B_i$) as well.

The CARSIM model is compatible with the Longitudinal Control Model. As a matter of fact, A3 is resulted when $\dot{x}_i$ is set to zero in Equation 29.1. A1 is obtained when the vehicle interaction term (i.e. F^j_i) becomes zero. As vehicle i speeds up, Equation 29.1 predicts that the actual acceleration decreases, which is reflected in Tables 1 and 2 in the original paper [9]. A3 is found when $\dot{x}_i$ is equal to v_i in Equation 29.1. 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.

Psycho-Physical Model

The model developed by [157] represents a typical psycho-physical model whose principle is depicted in Figure 28.2. The rough line ABCD is a vehicle trajectory in the $\Delta x - \Delta \dot{x}$ plane where Δx is the spacing between the subject vehicle i and its leader j, i.e., $\Delta x = s_{ij}$, 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 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 vehicle trajectory intersects the 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 becomes sufficiently close and their speed difference is small. After point C, the two vehicles are in car following. As driver i tries to adapt to vehicle j's speed, the gap closes up. Driver i stops deceleration when the two vehicles are moving at the same speed and their distance remains constant. Considering that driver i may temporarily lose attention (e.g. on cell phone) and slow down. In this case, the gap begins to open till the driver realizes that he or she is falling behind. Consequently, the driver tries to catch up and, hence, the gap closes up 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 $\Delta x - \Delta \dot{x}$ plane.

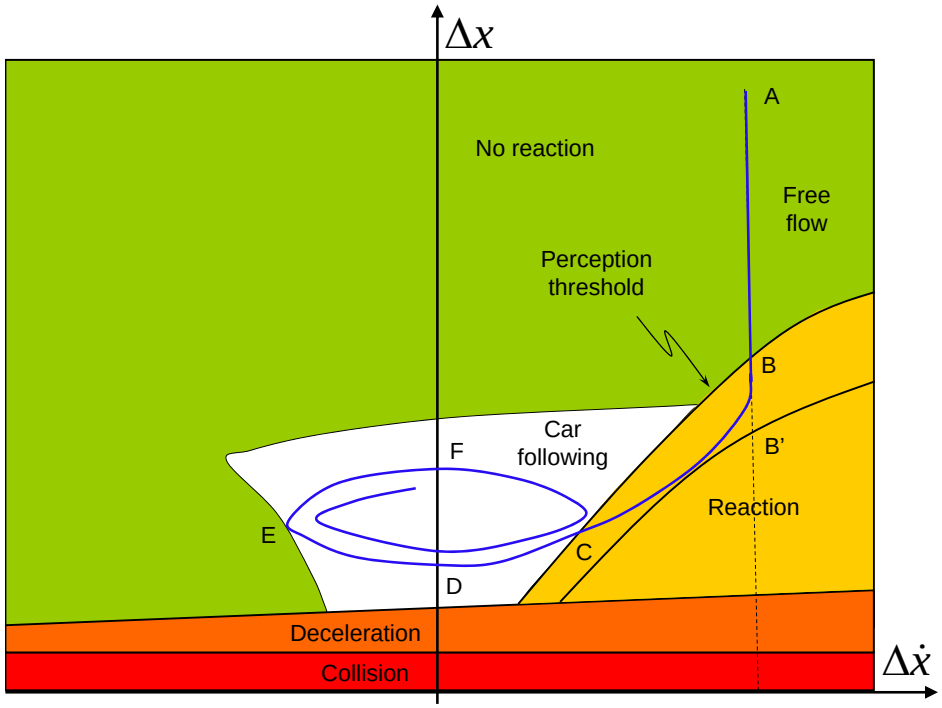


Figure 28.2: The Wiedmann model

Though the psycho-physical model is not directly contained in or derived from the Field Theory like the above models, 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 repulsion 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 will find an equilibrium spot 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 spot unconsciously, as predicted by Equation 28.6.

Rule-Based Model

A representative of rule-based models is the one developed by [76] which is reproduced below:

1. NO SPEED CHANGE

   Keep the present speed level (default case).

2. ACCELERATE IF [$\dot{x}_i < v_i$] and [$t - t_{last} > T_{acc}(\dot{x}_i)$]

   The current speed $\dot{x}_i$ is less than desired speed v_i and the time elapsed from last acceleration t_{last} is more than T_{acc}.

3. NO ACCELERATION IF [$s_{ij} < s_{min}(\dot{x}_i, \dot{x}_j) + w_{stab}(\dot{x}_i, \dot{x}_j)$]

   The distance from obstacle s_{ij} is less than the minimum safe distance s_{min} plus the width of stable area w_{stab}.

4. SLOW DOWN IF [$s_{ij} < s_{min}(\dot{x}_i, \dot{x}_j)$]

   The distance from obstacle Dobs is less than the minimum safe distance s_{min}.

5. DO NOT SLOW DOWN IF [$\dot{x}_i < \dot{x}_j$] or [$t - t_{last} < T_{maxdec}$]

   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).

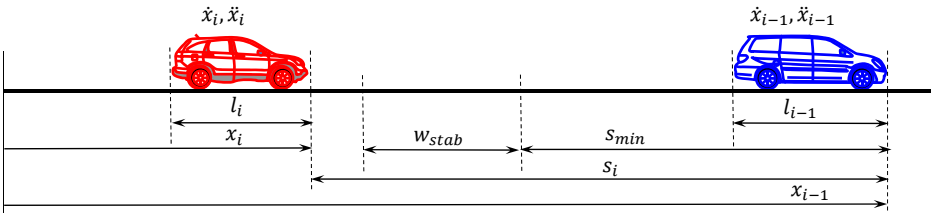


Figure 28.3: The Kosonen model

At each time step, the motion of i is checked against the above rules one by one. A latter rule always supersedes any earlier rules should there be a conflict. Similar to the situation in the psycho-physical 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 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 28.3, to mimic the original setup in [76]. 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 i's desire for mobility has not been fully satisfied (i.e., $\dot{x}_i < v_i$), it would accelerate (Rule 2). If i approaches j and is within w_{stab}, i would 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 would stop if it collides with j, which is ensured by the steep potential field when the vehicles touch (Rule 6).

28.2.3 Relating macroscopic equilibrium models

Under equilibrium condition, the following are noted: $\tau_i \to \tau$, $\dot{x}_i = \dot{x}_j \to v$, $v_i = v_j \to v_f$, $\ddot{x}_i = \ddot{x}_j = 0$, $s_{ij} \to s = \frac{1}{k}$, $s_{ij}^* \to s^* = \frac{1}{k^*}$, $b_i = b_j \to b$, $B_i = B_j \to B$, and $L_i = L_j \to \frac{1}{k_j}$ where the "$\to$" sign means "aggregate to" and items before the sign are microscopic variables and those after are macroscopic. With this notation, the microscopic Longitudinal Control Model (Equation 29.1) translates to its macroscopic counterpart (Equation 28.11 or more specifically Equations 29.3 and 29.4) 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)

Applying the above notations, Newell car-following model translates to its macroscopic counterpart of the following form:

$$v = v_f[1 - e^{\frac{\lambda}{v_f}\frac{1}{k_j}(1-\frac{k_j}{k})}] \tag{28.21}$$

Notice the close resemblance between Equations 29.10 and 28.11. In addition, through its microscopic counterpart, the above model's connection to the Longitudinal Control Model has been discussed in §28.2.2.

Van Aerde Model (Macroscopic)

The equilibrium counterpart of the Van Aerde Model can be written as:

$$k = \frac{1}{c_1 + c_3 v + c_2/(v_f - v)} \tag{28.22}$$

where all variables are defined before. Through its microscopic counterpart, the above model is connection to the Longitudinal Control Model as discussed in §28.2.2.

Intelligent Driver Model (Macroscopic)

Under equilibrium condition, a special macroscopic case was derived from the Intelligent Driver Model (IDM) [145, 55]:

$$v = \frac{(s-L)^2}{2v_f T^2}[-1 + \sqrt{1 + \frac{4T^2 v_f^2}{(s-L)^2}}] \tag{28.23}$$

where T is average safe time headway and $s = 1/k$ is average spacing and k traffic density.

Pipes-Munjal Model

Pipes-Munjal model [97] takes the following form:

$$v = v_f[1 - (\frac{k}{k_j})^n] \tag{28.24}$$

where n is a coefficient and other variables are defined before. Thinking in the reverse direction (i.e., from macroscopic to microscopic[1]), the above model seems to suggest a microscopic basis of roughly the following form:

$$\ddot{x}_i = g_i[1 - \frac{\dot{x}_i}{v_i} - (\frac{L}{s_{ij}})^n] \tag{28.25}$$

Note that the microscopic basis may take many other forms and the above represents only one of 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 $(\frac{L}{s_{ij}})^n$. 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

Drew model [31] takes the following form:

$$v = v_f[1 - (\frac{k}{k_j})^{n+\frac{1}{2}}] \tag{28.26}$$

where all variables are defined before. Repeating the above technique and replacing n with $n + \frac{1}{2}$, the suggested microscopic basis is:

$$\ddot{x}_i = g_i[1 - \frac{\dot{x}_i}{v_i} - (\frac{L}{s_{ij}})^{n+\frac{1}{2}}] \tag{28.27}$$

which can be derived from the Field Theory by choosing $F_i^j = g_i(\frac{L}{s_{ij}})^{n+\frac{1}{2}}$.

Wang Model

Wang model [151] recently proposed a stochastic equilibrium model whose three-parameter deterministic version takes the form of:

$$v = \frac{v_f}{1 + e^{\frac{k-k_c}{\theta}}} \tag{28.28}$$

where k_c is critical density (i.e. the density after which speed drop becomes noticeable as density increases from 0 to k_j) and θ a coefficient. The microscopic basis of the model could be:

[1]The same technique was used to derive Van Aerde car-following model (microscopic) from Greenshields model (macroscopic)

$$\ddot{x}_i = g_i[1 - \frac{\dot{x}_i}{v_i} - (1 - \frac{1}{1 + e^{\frac{1}{\theta}(\frac{1}{s_{ij}} - \frac{1}{s^c})}})] \tag{28.29}$$

where $s^c = 1/k_c$ is the critical spacing (i.e. average spacing at critical density). According to the Field Theory, one only need to choose $F_i^j = g_i[1 - 1/(1 + e^{\frac{1}{\theta}(\frac{1}{s_{ij}} - \frac{1}{s^c})})]$ to obtain Wang model.

Del Castillo Model

Del Castillo [23, 24] proposed a family of exponential generating functions which can be represented as

$$f(\lambda) = e^{1-(1+\frac{\lambda}{n})^n} \tag{28.30}$$

where λ is called "equivalent spacing" which is a function of density k and n is a parameter. Setting $n = 1$ and $n \to \infty$ results in the following two special cases, respectively:

$$v = v_f[1 - e^{\frac{|C_j|}{v_f}(1 - \frac{k_j}{k})}] \tag{28.31}$$

and

$$v = v_f[1 - e^{1 - e^{\frac{|C_j|}{v_f}(\frac{k_j}{k} - 1)}}] \tag{28.32}$$

where C_j is kinematic wave speed at jam density and other variables are defined before. Equation 28.32 is referred to as the "maximum sensitivity curve". Equation 29.11 takes a similar form to Newell model and the Longitudinal Control Model. If one chooses $|C_j| = \lambda/k_j$, Equation 29.11 becomes Newell model and hance is connected to the Longitudinal Control Model. If the conjecture that $\lambda = 1/\tau$ were 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 meters and τ around 1 second. This yields $|C_j|$ around 5-10 m/s or 11-22 mph which agrees well with the numbers provided in [23].

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 Newell model. In addition, the family of speed-density curves can be represented generically as

$$v = v_f[1 - e^{\psi(k)}] \tag{28.33}$$

where $\psi(k)$ is a generic function and admits the corresponding terms in Equations 29.11 and 28.32. From the perspective of the Field Theory, Del Castillo seems to suggest a vehicle interaction force proportional to: $F_i^j \propto e^{\psi(1/s_{ij})}$.

GM-Associated models

In addition, The family of equilibrium models, including [50], [49], [146], and [29] models, which are associated with GM models has been discussed in Chapter 19.

28.3 The Unified Diagram

Summarizing the discussion above, a unified perspective can be casted on these traffic flow models. Such a perspective is presented as a diagram in Figure 28.4.

28.3.1 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 both longitudinal x and lateral y directions on a planar surface. The Field Theory formulated in Equation 28.6 belongs to this category. The middle pane has *microscopic* car-following models which only describe vehicle motion in 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 Newell model. The right pane includes *macroscopic* models which describe equilibrium speed-density relationships. The connection 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

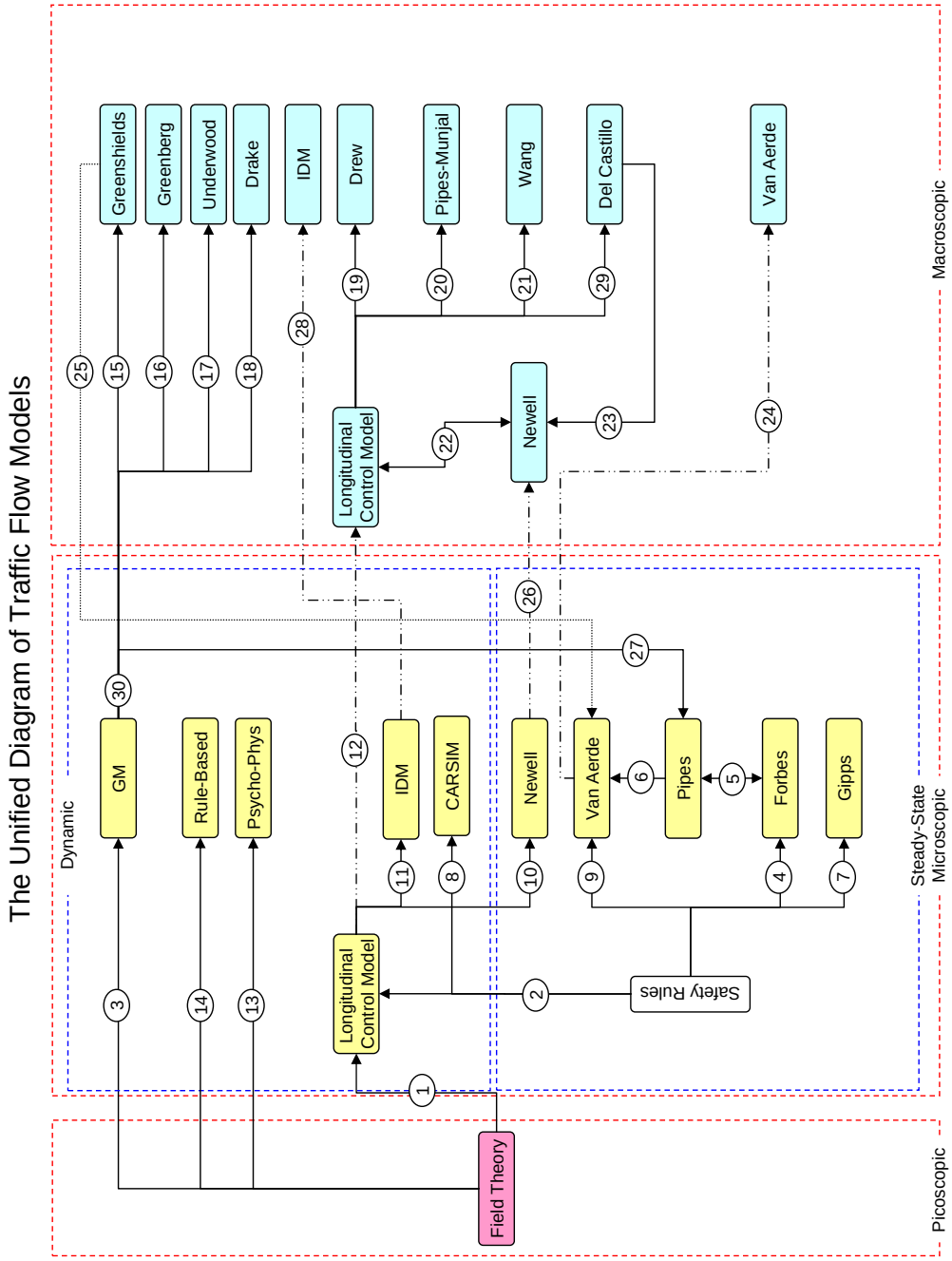


Figure 28.4: The United Diagram

the microscopic Longitudinal Control Model (Equation 29.1) and Newell nonlinear car-following model (Equation 29.8). By entering the list below, the reader is referred to §28.1 and §28.2.2 for the discussion on such a relation.

28.3.2 List of Connections in the Unified Diagram

This subsection references connection numbers in Figure 28.4 to the proper locations in the text where the nature of these connections are discussed.

- #1: See §28.2.1 for Longitudinal Control Model as a special case of the Field Theory.
- #2: See §28.2.1 for safety rules being admitted into the Longitudinal Control Model.
- #3: See §28.2.2 for for GM models as a special case of the Field Theory.
- #4: See §28.2.1 and §28.2.2 for Forbes model as a safety rule.
- #5: See Chapter 19 for the equivalence between Pipes model and Forbes model.
- #6: See §28.2.2 for Pipes model being admitted into Van Aerde model.
- #7: See §28.2.2 and §28.2.1 for Gipps model as a safety rule.
- #8: See §28.2.2 for the relation between CARSIM and the Longitudinal Control Model as well as the safety rule in CARSIM.
- #9: See §28.2.2 for Van Aerde model as a safety rule.
- #10: See §28.1 and §28.2.2 for Newell model as a special case of the Longitudinal Control Model.
- #11: See §28.2.2 for the relation between IDM and the Longitudinal Control Model.
- #12: See §28.2.1 for deriving macroscopic counterpart of the Longitudinal Control Model.
- #13: See §28.2.2 for the relation between Psycho-Physical model and the Longitudinal Control Model.

- #14: See §28.2.2 for the relation between Rule-Based model and the Longitudinal Control Model.
- #15: See Chapter 19 for Greenshields model being derived from GM5.
- #16: See Chapter 19 for Greenberg model being derived from GM5.
- #17: See Chapter 19 for Underwood model being derived from GM5.
- #18: See Chapter 19 for Drake model being derived from GM5.
- #19: See §28.2.3 for the the relation between Drew model and the Longitudinal Control Model.
- #20: See §28.2.3 for the the relation between Pipes-Munjal model and the Longitudinal Control Model.
- #21: See §28.2.3 for the the relation between Wang model and the Longitudinal Control Model.
- #22: See §28.2.3 for the close resemblance between Newell model (macroscopic) and the Longitudinal Control Model.
- #23: See §28.2.3 for the equivalence between Newell model (macroscopic) and one of the special cases derived from Del Castillo's family of exponential generating functions.
- #24: Refer to [147, 148] for microscopic and macroscopic versions of Van Aerde model.
- #25: See §28.2.2 for Greenshields model being admitted into Van Aerde model.
- #26: See §28.2.3 for obtaining macroscopic counterpart of Newell car-following model.
- #27: See Chapter 19 for Pipes model being derived from GM5.
- #28: See §28.2.3 for macroscopic IDM being derived from its microscopic counterpart.
- #29: See §28.2.3 for how the Field Theory is related to Del Castillo model.

- #30: See Chapter 19 for how May's original unifying effort fits into the larger Unified Diagram.

28.4 Conclusion

Motivated by Newell's untold secret in his non-linear car-following model and May's original unifying effort depicted in Figure 6.6 of May [88], a broader unified perspective was casted on traffic flow modeling and a larger Unified Diagram was constructed.

Newell model [99], after being rearranged slightly, gives rise to a mechanics model which involves non-contact forces which, in turn, can be explained conveniently using the concept of field. The Field Theory of this nature has been proposed in Chapter 27 and is concisely reproduced here for easy reference. Using 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 varying 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 later. When aggregated, many of these car-following models reduce to their macroscopic counterparts, i.e. equilibrium speed-density relationships. Those macroscopic equilibrium models, which do not come with a proposed microscopic basis, fortunately contains information to deduce their microscopic nature, though which the connection with the Field Theory might be established.

Summarizing the above analysis, the Unified Diagram is constructed which gather together traffic flow models at the picoscopic, microscopic, and macroscopic levels with lines connecting related models. Each connection is denoted by a number referencing to the discussion of the nature of such a connection in the text.

Chapter 29

Validation and Benchmarking

29.1 Introduction

Over more than half a century, traffic flow theorists have been pursuing two goals: (1) simple and efficient models to abstract vehicular traffic flow and (2) a unified framework in which existing traffic flow models fit and relate to each other. In response to the second goal, a framework and, further, a Unified Diagram has been proposed in Chapter 28. In that chapter traffic flow models at both microscopic and macroscopic levels are directly or indirectly related to the Field Theory that is formulated Chapter 27.

The first goal gives rise to the issue of how to evaluate models. Unfortunately, a set of commonly agreed evaluation criteria has yet to be devised since this is a subjective matter. Nevertheless, the following proposition may be approved by many: *empirical* validity, *mathematical* elegance, and *physical* soundness, which are further elaborated as follows.

Empirical Validity

A model is empirically valid if it agrees well with observations in the laboratory and field.

Mathematical Elegance

A model is mathematically elegant if it summarizes the underlying process using the least number of equations (ideally one equation for all) in simple functional form involving the least number of parameters/coefficients.

Take car-following models for example, the General Motors (GM) models [11, 43] involve one equation, the [46] model consists of two equations, psychophysical models [157] features multiple equations, rule-based models [73, 76] utilizes a set of fuzzy decision rules rather than mathematical equations, and agent-based models [111] employ artificial intelligence (e.g. neural networks) to make decisions. As one moves along this list, the models become increasingly intelligent. Meanwhile, their mathematical tractability progressively diminishes.

Physical Soundness

A model is physically sound if it is able to capture the underlying mechanism of the target system in a way that is consistent with first principles (such as Newton's laws of motion or Maxwell's equations of electromagnetism). In general, there are two modeling approaches: *descriptive* vs. *explanatory*. The descriptive approach focuses on fitting the model to observed behavior of the target system without worrying about how the system works, while the explanatory approach looks behind the scene and tries to represent the underlying principle by following its chain of reasoning. For example, the [50] model stipulates that when traffic density is x traffic speed is y, but the model could not explain why this is the case. In contrast, the [46] model predicts that a driver moving at v km per hour is likely to leave a spacing of s m ahead. This is so because it is calculated from the driver's control strategy which guarantees his or her safety in case that the leader applies emergency brake. Though descriptive models may be simpler in some cases, explanatory models are generally more informative and traceable.

In addition, it is desirable that a microscopic model aggregates to known macroscopic behavior and a macroscopic model has sound microscopic basis. For example, a car-following model describes the behavior of individual vehicles. Aggregation of such behavior over vehicles and time gives rise to a macroscopic speed-density relationship. Hence, a sound car-following model would statistically agree with observed relation between traffic speed and density. Conversely, a sound macroscopic speed-density model should be rooted in or can be derived from a valid car-following model. Such a micro-macro bridge is attractive because it allows an additional, complementary perspective on the same system and thus it may be more efficient to identify and address operational deficiencies of the system.

Moreover, it is preferable that a model employs parameters that are phys-

ically meaningful and easy to calibrate. Model parameters can be roughly divided into the following three categories: (1) Parameters with physical meaning and easy to calibrate (e.g. involving a small-scale data collection effort, based on which determining the value of the parameter is a relatively easy call). Included in this category are desired speed, perception-reaction time, vehicle length, maximum acceleration, normal deceleration, maximum deceleration, free-flow speed, jam density, capacity, etc. (2) Parameters with physical meaning but difficult to calibrate (e.g. involving large-scale data collection efforts, based on which determining the value of the parameter may not be a clear call). (3) Coefficients that need calibration and without physical meaning.

With the above understanding, the objective of this chapter is to evaluate the Longitudinal Control Model as well as a subset of the models presented in the Unified Diagram, all of which meet the above criteria. This chapter is arranged as follows. First, the validity of the Longitudinal Control Model (LCM) will be assessed by comparing against empirical observations and the comparison will be conducted at both microscopic and macroscopic levels (Section 29.2). In addition, it is interesting to cross-compare the proposed model with some of the well-established models presented in the Unified Diagram. This will be accomplished by benchmarking these models against a common set of data (Section 29.3). Finally, concluding remarks will be presented in Section 29.4.

29.2 Validation

The Field Theory has been formulated in Chapter 27 with particular interest in its special case, the Longitudinal Control Model, which is reproduced below for easy reference.

Longitudinal Control Model (microscopic) This model describes the longitudinal control behavior of driver-vehicle unit i in response to its leading vehicle j:

$$\ddot{x}_i(t+\tau_i) = g_i[1 - (\frac{\dot{x}_i(t)}{v_i}) - e^{\frac{s_{ij}(t)^* - s_{ij}(t)}{s_{ij}(t)^*}}] \tag{29.1}$$

where $\ddot{x}_i(t)$ denotes vehicle i's acceleration at time t, τ_i perception-reaction time of driver i, $\dot{x}_i$ vehicle i's speed, g_i the maximum acceleration that driver

i is willing to apply when starting from stand still, v_i driver i's desired speed, s_{ij} the spacing between vehicles i and j, and s_{ij}^* the desired spacing between i and j which can be further determined as:

$$s_{ij}^*(t) = \frac{\dot{x}_i^2(t)}{2b_i} + \dot{x}_i(t)\tau_i - \frac{\dot{x}_j^2(t)}{2B_j} + L_j \tag{29.2}$$

where $b_i > 0$ driver i's deceleration rate, $B_j > 0$ driver j's maximum deceleration rate, and L_j vehicle j's minimum safety room or nominal vehicle length.

Longitudinal Control Model (macroscopic) Derived from its microscopic counterpart, this model describes the macroscopic equilibrium speed-density relationship:

$$v = v_f[1 - e^{1 - \frac{1}{k(v\tau' + L)}}] \tag{29.3}$$

Or

$$k = \frac{1}{(v\tau' + L)[1 - ln(1 - \frac{v}{v_f})]} \tag{29.4}$$

where driver's desired speed v_i aggregates to free flow speed v_f, i.e. $v_i \to v_f$; individual spacing s_{ij} aggregates to average spacing s which is the reciprocal of traffic density k, i.e. $s_{ij} \to s = \frac{1}{k}$; similarly $s_{ij}^* \to s^* = \frac{1}{k^*}$, where $s^* = \frac{v^2}{2b} + v\tau - \frac{v^2}{2B} + L$, b is average deceleration, B is average emergency deceleration, $L = \frac{1}{k_j}$ is average minimum spacing, and k_j is jam density. If one ignores the difference in braking (i.e. $b = B$) and considers the effect that drivers are more vigilant at higher speeds (i.e. replacing the constant perception-reaction time τ with a speed-varying one $\tau' = \tau e^{-v/v_f}$), the desired spacing reduces to $s^* = v\tau' + L$.

This section assesses the validity of the Longitudinal Control Model (both microscopic and macroscopic) against empirical observations. Two challenges are faced by model validation. The first is the availability of empirical data. Ideally, field experiments need to be devised in order to customize data collection to the specific needs of model validation. However, such experiments can be costly or prohibitive depending on the nature of these experiments. Hopefully, this challenge can be addressed by making use of data published

by authorities or obtained through open sources. In this study, the empirical data to be used include vehicle trajectories (3 data sets), aggregated traffic flow observations (6 data sets), road test results (2 sets), and regulation standard (1 set). The second is validation methodology. Normally, statistical methods are employed to compare model output against the corresponding real world process, both of which involve randomness. To do so, a large amount of samples are collected from the model and the real system so that statistical analysis can be performed at the aggregated level. This will be addressed in Subsection 29.2.2. At the individual level, it is meaningless to compare a simulated run against a particular real-world process since the latter is affected by many random factors, a good portion of which can be neither calibrated nor incorporated in the model. However, the flip side (i.e. whether the model is able to reasonably reproduce an observed process) may cast some insight on the validity of the model. This will be addressed in Subsection 29.2.1.

29.2.1 Microscopic Validation

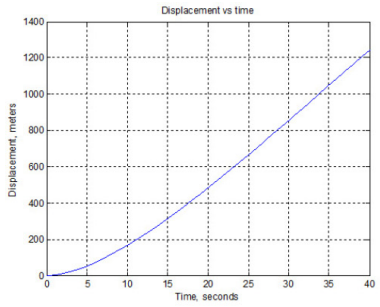
Validation at the microscopic level focuses on comparing model output against observations of individual driver-vehicle behavior. The comparison is conducted on acceleration, deceleration, and car-following performances.

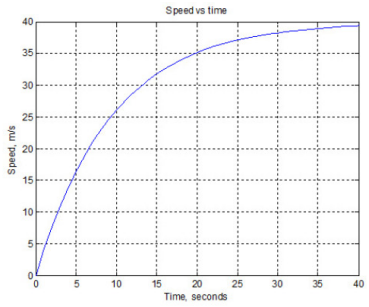
Acceleration Performance

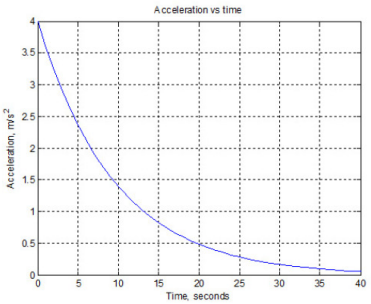
Figure 29.1 shows the acceleration performance of the Longitudinal Control Model which accelerates a vehicle from stand still to a pre-determined desired speed without being impeded by other vehicles. The simulation is set up as follows: $g_i = 4$ m/s^2, $v_i = 40$ m/s, $\tau_i = 1$ s, $b_i = 4$ m/s^2, $B_j = 6$ m/s^2, and $s_{ij} = \infty$. The top-left figure is displacement vs. time plot, the top-right speed vs time curve, bottom-left acceleration vs. time plot, and the bottom-right tabulates part of the simulation data (left 4 columns).

According to Motor Trend[1] road tests, it takes about 10.4 seconds to accelerate a passenger car from 0 to 26.8 m/s (60 mph) and 17.8 seconds to traverse a quarter mile (about 402 m). The Longitudinal Control Model predicts that 0-60 mph time is about 10.5 seconds and quarter mile time is 17.7 seconds. The two sets of results agree well.

[1] http://www.motortrend.com







Time (s)	x (m)	v (m/s)	a (m/s^2)	v' (m/s)	a' (m/s^2)
2	11.6	7.6	3.2		
4	36.2	13.8	2.6	13.4	2.7
6	71.3	18.7	2.1	17.9	2.2
8	115.0	22.8	1.7	22.4	1.8
10	165.5	26.0	1.4	26.8	1.6
11	193.0	27.4	1.3		
15	314.1	31.8	0.8		
18	413.9	34.0	0.6		
20	483.6	35.1	0.5		
25	665.5	37.1	0.3		
30	854.6	38.2	0.2		
35	1048.0	38.9	0.1		
40	1243.9	39.3	0.1		

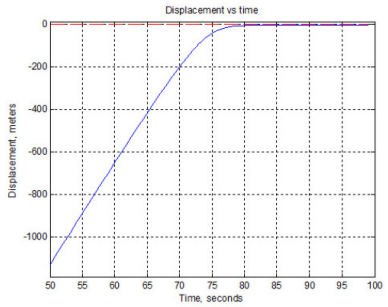
Figure 29.1: Acceleration performance

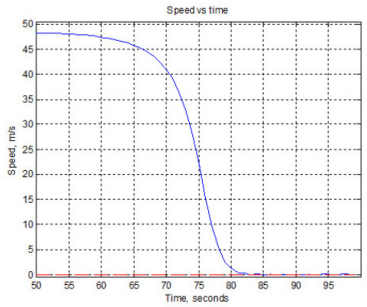
In addition, the effect of vehicle speed on acceleration published in NCHRP Report 185 [139] is presented in the last two columns of the table in Figure 29.1. For example, when speed is 13.4 m/s (30 mph), acceleration is 2.7 m/s^2. A comparison of the numbers show that the acceleration performance of the Longitudinal Control Model agrees roughly with those in the report.

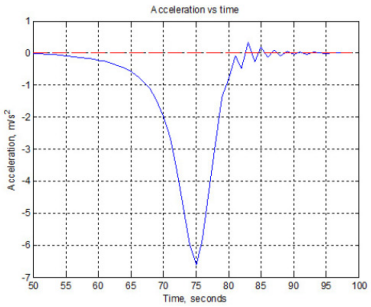
Deceleration Performance

Figure 29.2 shows the deceleration performance of the Longitudinal Control Model which brings a vehicle to a stop behind a stationary object (a stop line or a stopped vehicle). The simulation is set up as follows: initial speed $\dot{x}_i = 48$ m/s, initial position $x_i = -1127$ m, $x_j = 0$ m constantly, $g_i = 4$ m/s^2, $v_i = 50$ m/s, $\tau_i = 1$ s, $b_i = 4$ m/s^2, $B_j = 6$ m/s^2. The top-left figure is displacement vs. time plot, the top-right speed vs time curve, bottom-left acceleration vs. time plot, and the bottom-right tabulates part of the simulation data (left three columns). The model performance is compared against Part 571 of the Federal Motor Vehicle Safety Standards - Standard No. 105 [36] published by the Federal Motor Carrier Safety Administration (FMCSA). The purpose of the standard is to ensure safe braking performance

under normal and emergency conditions. The FMCSA standard, which is listed in the last two columns of the table, stipulates the distances required to decelerate a vehicle to a stop from its initial speeds. For example, a passenger car should be able to come to rest from 26.8 m/s (60 mph) within 65.8 m (216 ft). A comparison of the numbers suggests that the acceleration performance of the Longitudinal Control Model agrees reasonably with those in the FMCSA standard.







x (m)	v (m/s)	a (m/s^2)	v' (m/s)	x' (m)
364.7	45.2	-0.7		
320.2	44.5	-0.9	44.7	205.1
276.6	43.6	-1.2		
234.2	42.4	-1.5		
193.2	40.9	-2.0		
154.2	39.0	-2.6		
117.9	36.3	-3.6	35.8	123.4
85.1	32.8	-4.7		
57.1	28.0	-6.0	26.8	65.8
35.1	22.0	-6.6		
19.6	15.4	-5.8	13.4	16.5
10.0	9.6	-4.3		
0.0	0.0	0.0		

Figure 29.2: Deceleration performance

Car-Following Performance

Ideally, the data suitable for evaluating car-following performance can be collected in a way similar to what General Motors researchers did [11] where a leading driver was instructed to carry out pre-determined maneuvers while a second driver followed using common sense. Constrained by resources, this is not an option for this study. Alternatively, vehicle trajectory data are sought to fulfill the need. Three sets of trajectory data are used, from which pairs of vehicles in car following are identified. For each pair, the leading vehicle's trajectory, speed, and acceleration are used as input, against which

the model performs car following. Then the performances of the leading vehicle (observed) and the following vehicle (both observed and simulated) are plotted on the same figure for comparison. The objective is to investigate whether the model, after being tuned over a reasonable range, is capable of reproducing the observed car-following behavior.

This study uses three sets of trajectory data involving a total of 7 car-following pairs. The NGSIM data set contains detailed vehicle trajectories collected on eastbound Interstate 80 for a period of approximately half an hour between 2:35 and 3:05 PM on Wednesday, December 3, 2003. The vehicle trajectories are screened and car-following pairs are identified, among which three pairs are randomly selected: 1388-1386 (leader-follower), 3011-3012, and 4022-2193. The JHK data set was collected in 1985 [137] and vehicle trajectories were extracted from aerophotos taken at one frame per second. After pre-processing the data and plotting vehicle trajectories, it was found that good car-following pairs suitable for the current study were limited, of which three pairs were selected: 935-936, 2666-2668, and 4042-4045. The IVBSS data set was collected recently under a cooperative agreement between the U.S. Department of Transportation and the University of Michigan Transportation Research Institute along with its partners. The authors only have access to a sample file of the data set. Hopefully, this sample file recorded an extended trip of a driver with rich and time-stamped information about the vehicle's status (GPS location, speed, acceleration, and other in-vehicle sensor signals) and its driving environment (particularly useful are the distance and speed relative to its leading vehicle).

Figure 29.3 shows the results of car-following pair 1388-1386 (leader-follower). The top-left figure is displacement vs. time plot, the top-right speed vs. time plot, and bottom-left acceleration vs. time plot. The leader (empirical) is illustrated using solid black lines, the follower (empirical) dash blue lines, and the follower (simulated) dotted red lines. The bottom-right shows errors/residues of displacement (solid black line, m), speed (dash blue line, m/s), and acceleration (dotted red line, m/s^2) vs. time. As discussed before, the key issue here is to examine the model's ability to reproduce the follower's behavior. Therefore, one would check how the simulated follower approximates the empirical follower, particularly in terms of speed. This is because the displacement curves are too close to call due to their large base numbers and the acceleration curves are prone to random effects and high-frequency variation. In contrast, the speed curves not only highlight the vehicles' performances on the right scale but also integrate out random

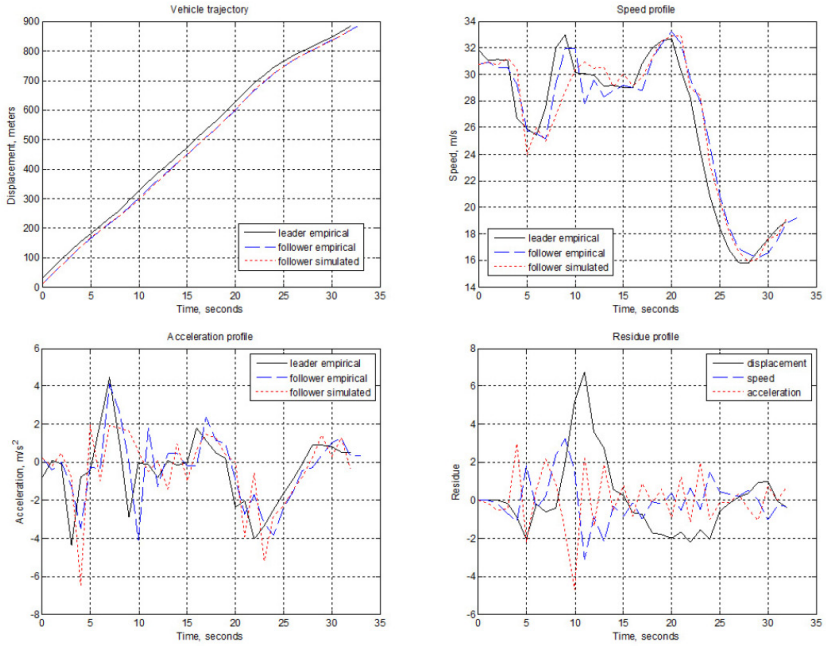


Figure 29.3: Car following performance (NGSIM I-80 data set, Leader: #1388, Follower: #1386)

effects and high-frequency variation. In addition to assessing the quality of approximation, one needs to verify if the simulated follower exhibits unacceptable behaviors such as crossing over its leader (which signifies a crash), moving backwards (which involves negative speeds), being out of pace with the empirical follower, particularly in the speed profile (which breaks car following), etc. With these in mind, the reader is entitled to make assessment on the model's performance. Comparison of other car-following pairs show similar results and are not plotted here due to space constraint. Values of parameters used in this microscopic car-following validation are tabulated in Figure 29.4. Apparently, these parameter values are within reasonable range. Though numerical comparison results may not as informative as those of graphical comparison, the former are tabulated in Figure 29.5 for complete information. For each car-following pair (e.g. "NGSIM 1388-1386"), the following are calculated: the mean ("Mean") and standard deviation ("STD") of modeling errors (i.e. $\{\varepsilon = y_{simulated} - y_{empirical}\colon\ y = (x, v,\ \text{or}\ a)\}$) of displacement ($x$), speed ($v$), and acceleration ($a$). Statistically, a zero mean of modeling errors indicates that the model is unbiased. However, this may result from large modeling errors with positive and negative errors canceling each other. Therefore, the second statistics, standard deviation, quantifies how much the simulated results deviate from the empirical results and an ideal fit would result in zero standard deviation. Hence, the mean and standard deviation combined can be an indicator of goodness of fit. The results in the table appear quite satisfactory.

	τ_i (s)	V_i (m/s)	L_j (m)	g_i (m/s^2)	b_i (m/s^2)	B_j (m/s^2)
1388-1386	0.3	40	4.3	6	6	6
3011-3012	0.4	40	4.3	6	6	6
4022-2193	0.6	41	4.3	6	6	6
935-936	0.5	34	4.6	6	6	6
2666-2668	0.7	32	4.6	6	6	6
4042-4045	1.5	28	4.6	6	6	6
IVBSS	0.4	36	5	6	6	6

Figure 29.4: Car following validation parameters

29.2.2 Macroscopic Validation

At the macroscopic level, the validation is conducted by comparing the performance of the model and its target real world system statistically by aggregating over driver population and time. The desirable empirical data are

		x (m)	v (m/s)	a (m/s^2)
NGSIM	Mean	-0.04	-0.01	0.06
1388-1386	STD	2.26	1.19	1.32
NGSIM	Mean	-0.04	-0.01	0.06
3011-3012	STD	2.57	1.15	1.23
NGSIM	Mean	0.04	-0.20	-0.02
4022-2193	STD	4.49	1.22	1.25
JHK	Mean	0.00	-0.09	0.01
935-936	STD	1.00	1.44	2.51
JHK	Mean	-0.02	-0.02	-0.06
2666-2668	STD	0.98	1.28	2.26
JHK	Mean	0.06	0.25	0.04
4042-4045	STD	3.15	1.17	1.80
IVBSS	Mean	1.42	-0.02	-0.01
	STD	8.64	1.19	1.03

Figure 29.5: Car following validation results

typically traffic flow characteristics such as traffic speed, density, and flow observed in the system. The desirable simulated data can be obtained by either conducting multiple runs of microscopic simulation with random parameters or reducing the microscopic model to a macroscopic one based on which the simulation is conducted. Since the macroscopic version of the Longitudinal Control Model (Equation 29.3 or 29.4) is readily available, the latter approach will be adopted. As such, the comparison will be performed between empirical traffic flow data and the macroscopic Longitudinal Control Model. The macroscopic model involves a few parameters, more specifically free-flow speed v_f, average perception-reaction time τ, and jam density k_j which are calibrated as follows. Free-flow speed v_f can be easily inferred from the empirical data, e.g. by averaging traffic speed observations when density is virtually zero. Jam density k_j can be inferred from empirical density observations when traffic speed is virtually zero, though such observations are typically rare. Alternatively, bumper-to-bumper distances between vehicles may provide a good basis to deduce jam density. The average perception-reaction time τ is typically not included in traffic flow observations. An ideal approach can be a human factors study involving a representative sample of the driving population. Alternatively, one may infer τ using observed flow as a basis, though this approach is less than ideal and might necessitate subjective judgements.

The data used in this validation are collected at multiple locations including: Georgia 400 in Atlanta (Stations GA4001116, GA4001120, GA4001139), Autobahn in Germany, I-4 in Florida, Highway 401, Amsterdam Ring Road,

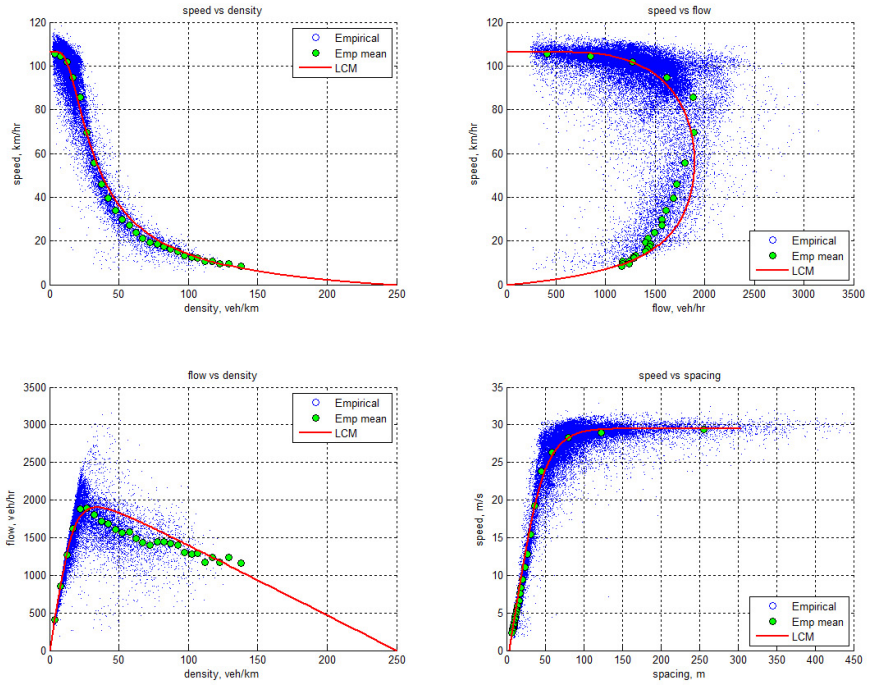


Figure 29.6: Comparison of fundamental diagrams (GA4001116 data)

and I-80 in California. For each set of data, the comparison of empirical observations and model performance are presented in their underlying fundamental diagrams (a fundamental diagram is a set of graphics including speed-density, speed-flow, flow-density, and speed-spacing plots). For example, Figure 29.6 presents the empirical and simulated fundamental diagrams based on GA4001116 data. The top-left part is speed-density plot, the top-right speed-flow plot, the bottom-left flow-density plot, and the bottom-right speed-spacing plot. In each plot, empirical observations ("Empirical") are scattered as small dots ("the cloud") and the simulated data ("LCM") are shown as solid lines. In addition, empirical observations are averaged with respect to density and the results ("Emp mean") are displayed as large dots. With the four plots of the fundamental diagram, the figure provides the full perspective needed to assess the overall performance of a macroscopic equilibrium model from complementary angles. In each data set, the speed-density

and speed-spacing plots appear to show encouraging results, while the flow-density and speed-flow plots (particularly the latter) tend to reveal problems. In the speed-flow plot, the empirical data seem to suggest an almost flat upper branch, i.e. the shape has a pointed "nose" that leans upward, which gives rise to a high optimal speed and a low optimal density at capacity. However, the model only exhibits a round nose, though it does lean upward in some cases, this is not enough. In addition, the model extrapolates a jam density of about 250 veh/km in some cases which translates to about an average inter-vehicle spacing of 4 m which is quite small unless the vehicles are predominantly compact cars with regular commuters[2]. Except for these, the model in general agrees well with the empirical data, especially in good match with the aggregated observations ("Emp mean").

Additional comparison based on the remaining data sets show similar results and are not plotted here due to space constraint. Values of model parameters used in this study are tabulated in Figure 29.7 and numerical comparison results including values of optimal density (k_m) and optimal speed (v_m) at capacity (q_m) are tabulated in Figure 29.8 for all data sets. Both visual and numerical comparison results suggest that the model agrees with empirical observations reasonably well.

	v_f (km/hr)	k_j (v/km)	tau (s)
GA4001116	106.2	250.0	1.4
GA4001120	104.4	250.0	1.2
GA4001139	104.4	200.0	1.2
Autobahn	155.9	83.3	0.8
I-4	88.2	83.3	0.4
Hwy401	106.2	83.3	0.6
Amsterdam	102.2	90.9	0.2
I-80	111.6	158.7	2.3

Figure 29.7: Macroscopic validation parameters

29.3 Benchmarking

At this point, the Field Theory and its special case, the Longitudinal Control Model, have been formulated and supporting evidences are presented. Next, it might be interesting to cross-compare this model and other well-established traffic flow models. Our goal is neither promoting nor diminishing certain

[2]This observation was prompted by Dr. Martin Treiber's work

			LCM			Empirical		
Data set	v_f (km/hr)	k_j (v/km)	q_m (v/hr)	v_m (km/hr)	k_m (v/km)	q_m (v/hr)	v_m (km/hr)	k_m (v/km)
GA4001116	106.2	250.0	1900.0	56.0	33.9	1887.0	69.4	27.2
GA4001120	104.4	250.0	2123.0	56.6	37.5	2145.0	78.5	27.3
GA4001139	104.4	200.0	1990.0	58.0	34.3	1978.0	72.2	27.4
Autobahn	155.9	83.3	2085.6	95.4	21.9	2114.0	95.0	22.3
I-4	88.2	83.3	1825.0	58.0	31.5	1785.0	70.0	25.5
Hwy401	106.2	83.3	1870.6	68.0	27.5	1946.0	89.1	21.8
Amsterdam	102.2	90.9	2509.0	68.0	36.9	2452.0	90.3	27.2
I-80	111.6	158.7	1181.0	58.3	20.3	1125.0	102.5	11.0

Figure 29.8: Macroscopic validation numerical results

models, but to provide a glimpse of the diversity of modeling flavors and how they complement each other that overall benefits the entire community. To accomplish this goal, we set up a common ground, on which each of the candidate models demonstrates so that their performances can be related to each other by referencing to the common ground. Such a process is called benchmarking and the common ground has been established in Section 16.3 involving a microscopic benchmarking scenario and macroscopic benchmarking scenario.

29.3.1 Microscopic Benchmarking

Model Selection

A subset of the microscopic car-following models presented in the Unified Diagram were selected including: General Motors 4th generation model (GM4) [11, 43], Gipps car-following model [46], Intelligent Driver Model (IDM) [145, 55], Newell non-linear car-following model [99], and Van Aerde car-following model [147, 148]. The selection mainly considers those models with simple, closed forms to facilitate implementation and comparison. For easy reference, the selected models are listed below and the reader is referred to the original papers for more details.

General Motors 4th generation model:

$$\ddot{x}_i(t+\tau_i) = \alpha \frac{\dot{x}_i(t+\tau_i)[\dot{x}_j(t) - \dot{x}_i(t)]}{[x_j(t) - x_i(t)]} \tag{29.5}$$

where x_i, $\dot{x}_i$, $\ddot{x}_i$, τ_i are the displacement, speed, acceleration, and perception-reaction time of the subject vehicle i respectively; similar notations apply to its leader j; α is a dimensionless coefficient.

Gipps car-following model:

$$\dot{x}_i(t+\tau_i) = \min \begin{cases} \dot{x}_i(t) + 2.5 g_i \tau_i (1 - \frac{\dot{x}_i(t)}{v_i}) \sqrt{0.025 + \frac{\dot{x}_i(t)}{v_i}} \\ b_i \tau_i + \sqrt{(b_i \tau_i)^2 + b_i [2(x_j(t) - L_j - x_i(t)) + \tau_i \dot{x}_i(t) + \frac{\dot{x}_j^2}{2B_j}]} \end{cases} \tag{29.6}$$

where g_i is the maximum acceleration that driver i wishes to apply, v_i is driver i's desired speed, $b_i > 0$ driver i's deceleration rate, B_j driver j's emergency deceleration rate, L_j vehicle j's nominal length.

Intelligent Driver Model:

$$\ddot{x}_i(t+\tau_i) = g_i [1 - (\frac{\dot{x}_i}{v_i})^\delta - (\frac{s_{ij}^*}{s_{ij}})^2] \tag{29.7}$$

where δ is acceleration exponent, $s_{ij} = x_j - x_i$ is the spacing between vehicle i and its leader j, and 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, and T_i are parameters.

Newell nonlinear car-following model:

$$\dot{x}_i(t+\tau_i) = v_i (1 - e^{-\frac{\lambda_i}{v_i}(s_{ij}(t) - L_i)}) \tag{29.8}$$

where λ_i is a parameter associated with driver i (i.e. the slope of i's speed-spacing curve evaluated at $\dot{x}_i = 0$) and L_i is the nominal length of vehicle i (also the minimum value of s_{ij}).

Van Aerde car-following model:

$$\dot{x}_i(t+\tau_i) = \frac{-c_1 + c_3 v_f + s_{ij}(t) - \sqrt{[c_1 - c_3 v_f - s_{ij}(t)]^2 - 4 c_3 [s_{ij}(t) v_f - c_1 v_f - c_2]}}{2 c_3} \tag{29.9}$$

where $c_1 = v_f(2v_m - v_f)/(k_j v_m^2)$, $c_2 = v_f(v_f - v_m)^2/(k_j v_m^2)$, $c_3 = 1/q_m - v_f/(k_j v_m^2)$, v_f is the free-flow speed of the roadway facility, k_j is the jam density, and v_m is the optimal speed occurred at capacity q_m.

Benchmarking Results

When determining values of model parameters, those suggested in the original papers are adopted by default. However, if such an adoption leads to inferior performance, appropriate new values will replace the default ones with good faith. Benchmarking results are plotted in Figure 29.9 which shows displacement vs. time, speed vs. time, and acceleration vs. time plots (from top to bottom, respectively). In each plot, the performance of the leader j is pre-determined and depicted as the black solid line, the performance of the follower i simulated by the Longitudinal Control Model is shown as the blue dashed line, and those of vehicle i simulated by the other models (i.e., GM, Gipps, Van Aerde, Newell, and IDM) are illustrated using different lines (differing by line type, color, and width). Values of model parameters used in this microscopic benchmarking is tabulated in Figure 29.10. Comments on benchmarking results are tabulated in Figure 29.11 and further elaborated as follows.

- L1: depending on parameter values, the model may have a little overshooting ($\ddot{x}_i(100) \approx -6.6$ m/s^2) in response to cutting off.
- L2: depending on parameter values, the model may have a little overshooting right before complete stop.
- M1: unable to start up from stand still. Need initial speed $\dot{x}(0) > 0$ or external logic.
- M2: acceleration profile does not make physical sense, i.e. acceleration is maximum when starting up, decreases as speed increases, and vanishes as desired speed is achieved.
- M3: there is no fixed free flow speed. The speed of vehicle i during free flow depends on its state when it was attracted or repelled by its leader last time.
- M4: rather than decelerating and backing up in response to cutting off, the model may react by acceleration.
- M5: the model may allow the vehicle to follow its leader arbitrarily close as long as the two vehicles are traveling at the same speed.
- M6: the model could not stop the vehicle completely. Without external logic, an infinitesimal speed has to be maintained in order to start the vehicle again. Hence, the vehicle may gradually crash into its stopped leader. Once the leader departs, the model needs an excessive time to warm up.

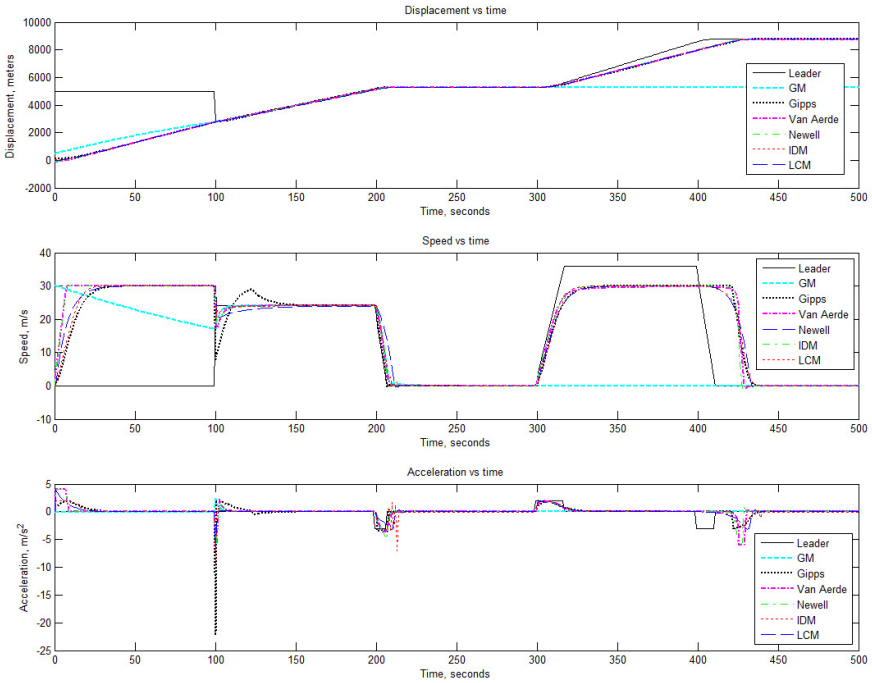


Figure 29.9: Microscopic benchmarking

- M7: the model predicts that the vehicle will be attracted by its speeding leader and will eventually adopt the latter's speed given sufficient time to warm up.
- G1: the model may involve oscillation during cutting off.
- G2: the model is conservative in car following under the original parameter setting
- G3 the model may under shoot during stopping
- D1: speed-up is slow under original parameter setting[3].
- D2: a little over-shooting during cutting off under original parameter setting

[3]Per suggestion of the author of the model, a higher acceleration parameter such as 2 m/s^2 to 2.5 m/s^2 and a higher desired speed such as 38 m/s might be more appropriate in this case

Model	τ_i (s)	v_i (m/s)	L_j (m)	g_i (m/s^2)	b_i (m/s^2)	More	
LCM	1	30	5	4	4	B_j = 6 m/s^2	
GM4	1	-	-	-	-	α = 0.8	
Gipps	1.1	30	6.5	1.7	3.4	B_j = 3.2 m/s^2	
IDM	T=1.6	30	5	0.73	1.67	δ = 4	s_0 = 2
Newell	1	30	5	-	-	λ = 0.79 1/s	
Van Ader	1	30			v_c = 25 m/s	q_c = 2/3 v/s	k_j = 143 v/km

Figure 29.10: Microscopic benchmarking parameters

Model	Start-up	Speed-up	Free flow	Cutting off	Following	Stop-n-go	Trailing	Approaching	Stopping
LCM	√	√	√	L1	√	√	√	√	L2
GM4	M1	M2	M3	M4	M5	M6	M7	√	M6
Gipps	√	√	√	G1	G2	√	√	√	G3
IDM	√	D1	√	D2	√	√	D1	√	√
Newell	N1	N2	√	N3	√	√	√	√	N4
Van Ader	V1	V2	√	V3	√	√	√	√	√

√ denotes satisfactory performance

Figure 29.11: Microscopic benchmarking results

- N1: Extreme start-up acceleration. Need external logic to constrain maximum acceleration.
- N2: Need another external logic to constrain the performance of acceleration on speed.
- N3: May over shoot depending on parameter setting
- N4: May over shoot depending on parameter setting
- V1: same as N1.
- V2: same as N2.
- V3: same as N3

These models can be divided into two groups: dynamic models (LCM, GM4, and IDM) and steady-state models (Gipps, Newell, and Van Aerde). A known issue with steady-state models is the extreme acceleration at start-up, i.e. the model allows a vehicle's speed to increase from zero to an arbitrary high speed (e.g. 30 m/s) resulting an unrealistic acceleration (e.g. 30 m/s^2). Hence, an external logic has to be applied to constrain the maximum acceleration. In addition, a another external logic is needed to constrain the performance of acceleration on speed, i.e. as speed increases, acceleration decreases till it vanishes at the achievement of desired speed. Gipps model is an exception since it trades model elegance (i.e. its two-piece formulation) for realistic acceleration performance. Overall, these models perform very well except for the known issues with GM model (nevertheless, it necessitates

only two parameters), especially considering that the series of challenging tests are faced by models with such simple and elegant forms.

29.3.2 Macroscopic Benchmarking

Model Selection

A subset of the macroscopic equilibrium models presented in the Unified Diagram were selected including: Newell model (macroscopic) [99], Del Castillo model [23, 24], Intelligent Driver Model (IDM, macroscopic) [145, 55], and Van Aerde model (macroscopic) [147, 148]. The preference was given to (1) single regime models with closed form, (2) models with sound microscopic basis (i.e. preferably derived from their corresponding car-following models), and (3) models with known satisfactory performance. For easy reference, the selected models are listed below and the reader is referred to the original papers for more details.

Newell model (macroscopic)

$$v = v_f[1 - e^{\frac{\lambda}{v_f}\frac{1}{k_j}(1-\frac{k_j}{k})}] \tag{29.10}$$

where v is traffic speed, k traffic density, v_f free-flow speed, λ a coefficient, and k_j jam density.

Del Castillo model

$$v = v_f[1 - e^{1-e^{\frac{|C_j|}{v_f}(\frac{k_j}{k}-1)}}] \tag{29.11}$$

where C_j is kinematic wave speed at jam density. This is referred to as the "maximum sensitivity curve" which is a special case derived from a family of exponential generating functions.

Intelligent Driver Model (IDM, macroscopic)

$$k = \frac{1}{(s_0 + vT)[1 - (\frac{v}{v_f})^{\delta}]^{-1/2}} \tag{29.12}$$

where s_0 is minimum gap, T is average safe time headway and δ is a calibration coefficient.

Van Aerde Model (macroscopic)

$$k = \frac{1}{c_1 + c_3 v + c_2/(v_f - v)} \tag{29.13}$$

where all variables are defined before.

Benchmarking Results

When fitting models to the data, free-flow speed is first determined since this information is readily available in the data. Then jam density is determined by following the trend of the congested portion of the flow-density plot. With these two parameters set, other parameters are fine-tuned to achieve satisfactory results. Though each model is fitted to the data with good faith, there is no guarantee that the fitting results are optimal. Nevertheless, the results are able to reveal some critical properties of these models. Fitted models are plotted in Figure 29.12 along with empirical data (aggregated with respect to density). Values of model parameters are tabulated in Figure 29.13 and benchmarking results are summarized in Figure 29.14.

Overall, these models demonstrates very satisfactory performance, especially considering their simple elegant forms, a variety of regimes over the full density range, and the quality of fitting. For example, all of them have clearly defined free-flow speed v_f and jam density k_j which agree with the data well. All of the flow-density curves exhibit a concave shape and match the capacity q_m well. Their discrepancies, in a picky eye, are mainly related to the capacity condition (i.e. k_m and v_m when capacity occurs). In the speed-flow plot, the capacity is illustrated as the "nose" of each curve. Newell model has its nose pointing downward which is less than ideal, the noses of Del Castillo model and LCM point forward which is quite close (Del Castillo model appears even closer than LCM), and the noses of Van Aerde model and IDM lean upward which is desirable (see figures in Subsection 29.2.2 (those with the "cloud") for a confirmation of such a shape). These discrepancies give rise to the effects that the "heads" of the flow-density curves are skewed toward the left in varying degrees, and that the "corners" of the speed-spacing curves are sharpened to varying extent. However, on a positive note, Newell model, Del Castillo model, and LCM necessitate only three parameters with Newell model featuring a particularly simple, closed form of $v(k)$.

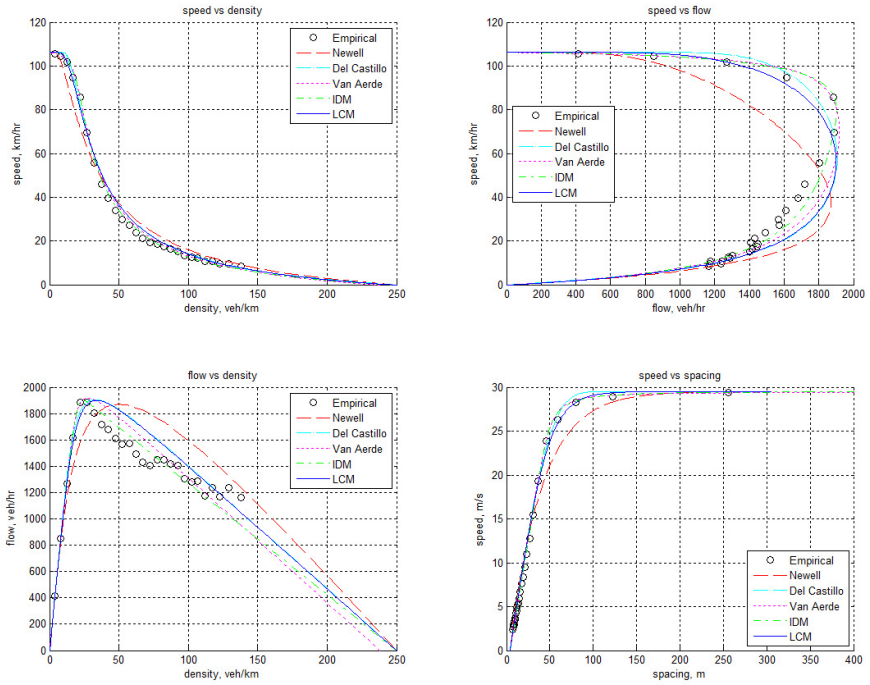


Figure 29.12: Macroscopic benchmarking (Data: GA4001116)

29.4 Conclusion

In an attempt to interpret highway traffic flow phenomena from a physical and social perspective, a Field Theory was put forth in Chapter 27. From the theory, a special case (i.e. the Longitudinal Control Model) was derived that is able to represent a driver's longitudinal control behavior at the microscopic level and the equilibrium traffic speed-density relationship at the macroscopic level. The microscopic Longitudinal Control Model was formulated in a simple equation with five parameters. The macroscopic Longitudinal Control Model, which is derived from its microscopic counterpart, is also represented by a simple equation with three parameters. All the parameters in both models have clear physical meaning and easy to calibrate. The focus of this chapter is to provide some empirical evidences in support of the Longitudinal Control Model as well as to cross-compare the performances of the model and

Model	v_f (km/h)	k_j (v/km)	More	
Newell	106.2	250	λ=0.79 1/s	
Del Castillo	106.2	250	$\|C_j\|$ = 2.6 m/s	
Van Ader	106.2	250	v_c = 72 km/r	q_c = 1950 v/h
IDM	106.2	250	t=1.7 s	δ = 15
LCM	106.2	250	τ=1.4 s	

Figure 29.13: Macroscopic benchmarking parameters (Data: GA4001116)

Model	v_f	k_j	v_m	k_m	q_m	λ	q-k shape
Newell	√	√	low	high	√	close	concave
Del Castillo	√	√	close	close	√	√	concave
Van Ader	√	√	√	√	√	√	concave
IDM	√	√	√	√	√	√	concave
LCM	√	√	close	close	√	√	concave

√ denotes satisfactory performance

Figure 29.14: Macroscopic benchmarking results (Data: GA4001116)

a set of well-established traffic flow models.

The validation of the Longitudinal Control Model was conducted at both microscopic and macroscopic levels. The microscopic validation compares the model's acceleration, deceleration, and car-follow performances against road test results, established standard, and empirical observations. Comparison results suggested that the model is able to realistically accelerate and decelerate a vehicle in a manner that complies with road test results and regulation standard when the vehicle is not impeded by other vehicles. When a leading vehicle is present, the validation translates to checking whether the model, after being tuned within reasonable range, is able to reproduce the behavior observed in the following vehicle. Three sets of observed vehicle trajectory data were used which involve three different locations and seven car-following pairs. Both graphical and numerical comparison results suggested that the model was indeed able to reproduce car-following behavior with reasonable accuracy. The macroscopic validation compares the model's macroscopic performance against empirical traffic flow data. Six sets of empirical data were used covering six different locations and 8 observation stations. The graphical comparison is based on the fundamental diagram which enables the full perspective to examine the observed and simulated speed-density, speed-flow, flow-density, and speed-spacing relationships. In each data set, the speed-density and speed-spacing plots appear to show encouraging results, while the flow-density and speed-flow plots (particularly the latter)

tend to reveal problems. In the speed-flow plot, the empirical data seem to suggest an almost flat upper branch, i.e. the shape has a pointed "nose" that leans upward, which gives rise to a high optimal speed and a low optimal density at capacity. However, the model only exhibits a round nose, though it does lean upward in some cases, this is not enough. In addition, the model extrapolates a jam density of about 250 veh/km in some cases which translates to about an average inter-vehicle spacing of 4 m which is quite small unless the vehicles are predominantly compact cars with regular commuters. Except for these, the model in general agrees well with the empirical data, especially in good match with the aggregated observations.

It is also interesting to cross-compare the Longitudinal Control Model with other traffic flow models of similar nature. a set of microscopic models and a set of macroscopic models are selected for this purpose. The preference of selection was given to (1) single regime models with closed form, (2) models with sound microscopic basis, and (3) models with known satisfactory performance. The cross-comparison employs the benchmarking technique by establishing a common ground, on which each model demonstrates its satisfactory performance and the results are related to those of other models. The microscopic benchmarking uses a hypothetical scenario as the common ground where multiple regimes encountered by a driver are integrated into one driving process. All the microscopic models performed well except for the known issues with the GM4 model and the acceleration issue with steady-state models (which can be quickly fixed with a constraint on limiting acceleration). The macroscopic benchmarking was based on one of the above empirical traffic flow data sets since everyone else appears equally good to serve this purpose. All the macroscopic equilibrium models are able to fit the empirical observations with reasonable accuracy except for some discrepancies under capacity conditions. More specifically, Van Aerde model and IDM fit the data best, and Del Castillo model and LCM are close to ideal. In addition, Newell model, Del Castillo model, and LCM necessitate only three parameters with Newell model featuring a particularly simple, closed form of $v(k)$.

Chapter 30

Multiscale Modeling of Traffic Flow

This chapter is a reproduction of the work in reference [107].

30.1 Introduction

Anyone who used maps probably developed the following experience. Fifteen years ago, a 1:10,000 paper map was needed to view a city (e.g. Amherst, MA), 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 overview the entire country, and then progressively zoom in to view Massachusetts, Amherst, and even the UMass Amherst campus, all seamlessly and within a single system.

Similarly, it is desirable that traffic simulation would allow an analyst to zoom in to examine low-level details and zoom out to overview system-wide performance within the same simulation process. Figure 30.1 illustrates such a paradigm. The background represents a *macroscopic* view of traffic operation in an entire region. This is analogous to viewing traffic 10,000 m above the ground and the traffic appears to be a compressible fluid whose states (speed, flow, and 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 3,000 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 is resulted. Similar to watching traffic

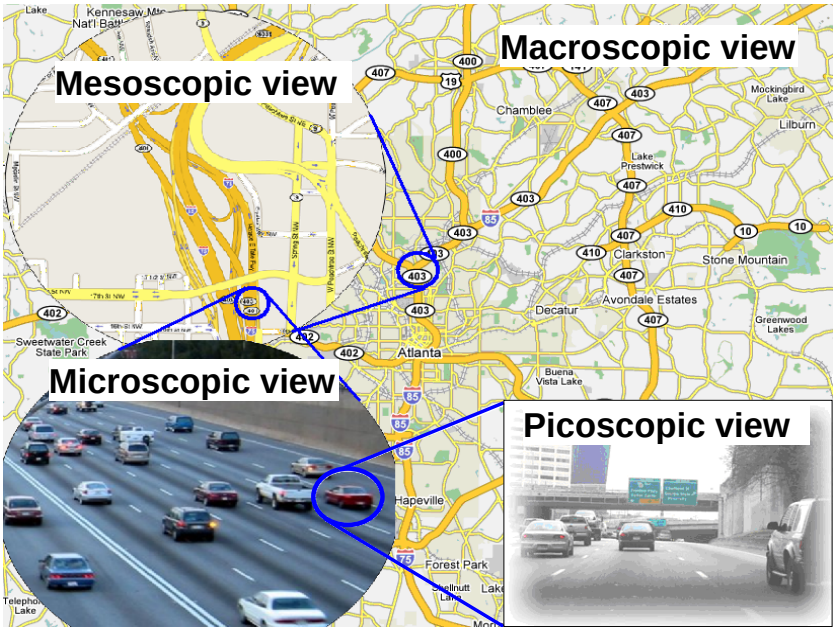


Figure 30.1: Multiscale traffic flow modeling

1,000 m above the ground, the scene is dominated by moving particles that interact with each other so as to maintain safe positions in 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, etc.), make control decisions, and manage vehicle dynamic respond to travel safely. If such a "zoomable" simulation becomes available, one would be able to translate traffic flow representation among multiple scales, e.g. to trace a low-level event all the way to a high-level representation and, conversely, to decompose a global problem down to 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, i.e. 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 30.2 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 30.3 presents the proposed multiscale approach based on field theory. Modeling strategy at each scale is discussed and some special cases are formulated at both the microscopic and macroscopic scales. The emphasis of this multiscale approach is to ensure coupling among different modeling scales. Concluding remarks and future directions are presented in Section 30.4.

30.2 The Spectrum of Modeling Scales

The modeling of traffic flow can be performed at, but is not limited to, a spectrum of four scales, namely picoscopic, microscopic, mesoscopic, and macroscopic from the most to the least detailed in that order. Considering that the definition 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 30.2 for each of the four modeling scales based on 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 the remaining three rows ("Underlying principle", "Modeling approach", and "Model coupling") pertain to the proposed multiscale approach with inherent consistency which are to be elaborated in the next section.

30.2.1 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 $i \in \{1, 2, 3, ..., I\}$ denotes vehicle ID, can be tracked in both longitudinal x and lateral y directions over time $t \geq 0$. Knowing these vehicle trajectories, the state and dynamics of the traffic system can be completely determined. 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

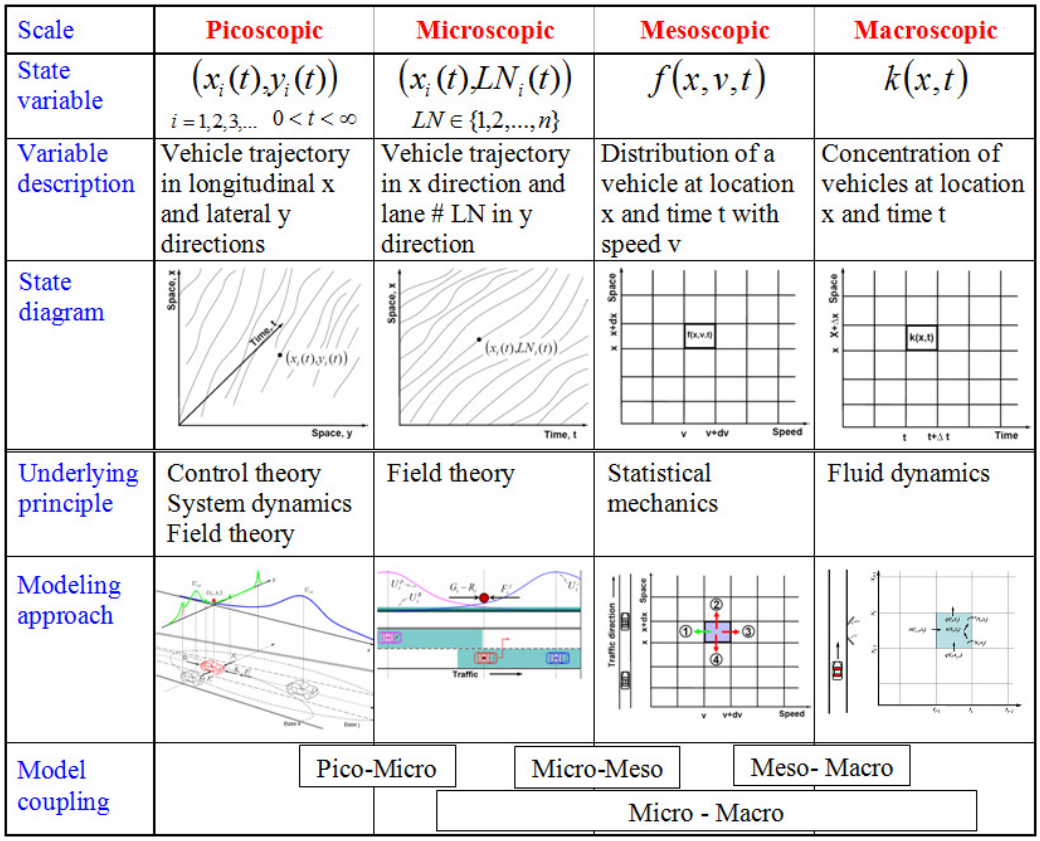


Figure 30.2: The spectrum of modeling scales

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 [3, 82]. 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 [158, 48]. 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 controllers. This observation allows the use of fuzzy logic [77, 41], which controls vehicles based on some predefined rules. To allow implicit driving rules, Artificial Neural Networks [85, 81] learn "driving experiences" from training processes and then apply the learned experiences in future driving. Several literature

surveys of driver models are available [162, 67, 84].

30.2.2 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 $LN_i(t)$ where $LN \in \{1, 2, ..., n\}$. Hence, $(x_i(t), LN_i(t))$ is state variable that describe the sate 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 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 [11, 43] 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 [119, 46] assume that the follower always attempts to achieve his desired gains (e.g. speed and safety), psycho-physical models [89, 157] introduce perception thresholds that trigger driver reactions, and rule-based models [76] 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 (MLC/DLC) [47, 4], adaptive acceleration MLC/DLC [62, 164], and autonomous vehicle control [140]. The following have been attempted to model gap acceptance: deterministic models [126, 149, 53], probabilistic models [59, 32, 60], and neuro-fuzzy hybrid models [133]. More surveys on microscopic models can be found in the literature[42, 64].

30.2.3 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 only of interest 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 $f(x, v, t)dxdv$ 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 $f(x, v, t)$, the dynamics of the system can be determined statistically.

Conventional mesoscopic traffic flow models come with three flavors. First, models such as the one in TRANSIMS [136] take a Cellular Automata approach where the space domain (representing the longitudinal direction of a highway) is partitioned in to short segments typically 7.5 meters long. If occupied, a segment is only able to store one vehicle. Vehicles are then modeled as hopping from one segment to another, so their movement and speed are discretized and can only take some predetermined values. Second, models such as those implemented in DynaMIT [8] and DYNASMART [12] use macroscopic models (such as speed-density relationship), as oppose to microscopic car-following models, to determine vehicle speed and movement. Third, truly mesoscopic models such as the one postulated by Prigogine and his co-workers [123] are based on non-equilibrium statistical mechanics or kinetic theory which draw analogy between classical particles and highway vehicles. Prigogine's model criticized [52] for (1) lacking 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 critiques 2 and 3. For example, Paveri-Fontana [113] considered a driver's desired speeds, Helbing [56] adapted the desired speeds to speed limits and road conditions, Phillips [117, 118] incorporated vehicle lengths, Nelson [98] accounted for vehicle acceleration behavior, and Klar and Wegener [155, 74] included a stochastic microscopic model. Surveys of previous approaches are available in the literature[57].

30.2.4 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 traffic condition. The state diagram typically involves a two dimensional domain (x, t). Knowing $k(x, t)$, the dynamics of the system can be determined 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[80, 130]. In addition, momentum and energy may also be conserved. A model is of a higher order if it incorporates the latter forms of conservation[114, 156]. Since the limited benefit offered by higher-order models often does not justify their added complexity[20], numerical approximation and macroscopic simulation have been centered on first-order models, e.g. KRONOS[91], KWaves[101], CTM[18, 21], FREQ[87], and CORQ [159]. More surveys of macroscopic models can be found in the literature[42].

30.2.5 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 following such an approach is able to support the "zoomzble" traffic simulation discussed in Section 30.1. 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 30.2. 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 using Edie's definition of traffic flow characteristics [33, 106].

30.3 The Proposed 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 comply to physical principles; the microscopic model

can be simplified from the picoscopic model yet still capturing the essential mechanisms of vehicle motion and interaction; the mesoscopic model can be derived from the microscopic model based on principles of non-equilibrium statistical mechanics; the macroscopic model can be derived from the mesoscopic model by applying principles of fluid dynamics. See a summary of underlying principle, modeling approach, and modeling coupling in Figure 30.2.

30.3.1 Picoscopic modeling

This section consolidates and highlights presentation in Chapters 24 and 27 as follows. In order 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. Based on principles of control theory, a driver-vehicle-environment closed-loop control (DVECLC) system has been developed. Figure 30.3 illustrates the components of the system and its control flow including feedback loops.

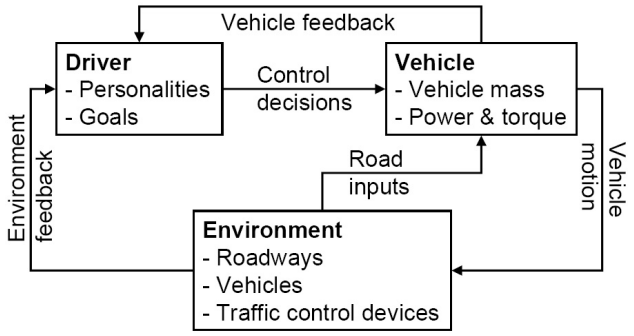


Figure 30.3: The closed-loop system

This system consists of a driver model and a vehicle model which interact with each other as well as with the driving environment. 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 rate. These sources of information, together with driver properties and goals, are used to determine driving strategies (such as steering and gas/brake). 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 constitute part of the environment. Other vehicle dynamic responses such as speed, acceleration, and yaw rate are fed back to the driver for determining driving strategies in the next step. Thus traffic operation is the collection of movement and interaction of all vehicles in the environment.

The driver model can be formulated by applying principles of field theory. Basically, objects in a traffic system (e.g. roadways, vehicles, and traffic control devices) are perceived by a 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. The superposition of these component fields represents the overall hazard encountered by the subject driver. Hence, the objective of vehicle motion 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.

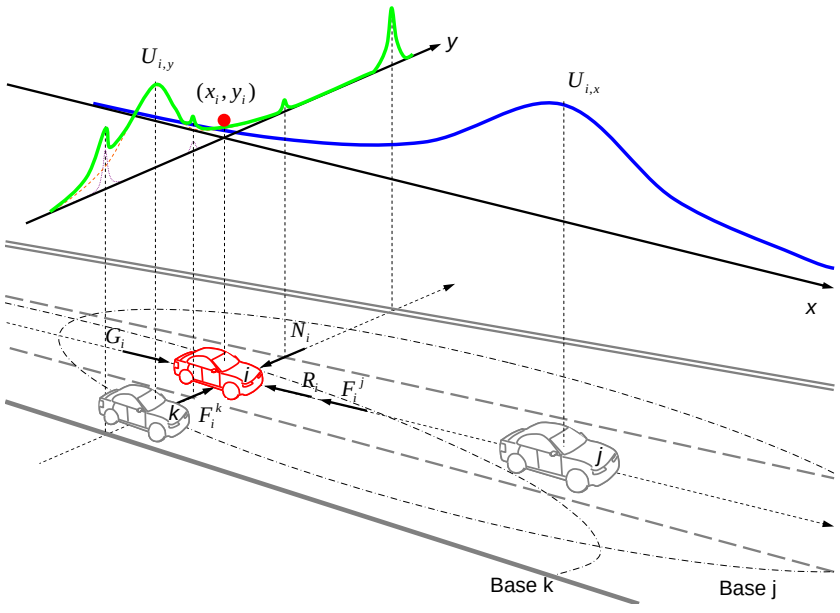


Figure 30.4: The illustration of a perceived field

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 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, i.e.

$$U_i = U_i^B + U_i^R + U_i^C$$

For example, Figure 30.4 illustrates two sections the overall field, $U_{i,x}$ and $U_{i,y}$. The subject 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:

$$F_i^j = -\frac{\partial U_{i,x}}{\partial x}$$

The effect of F_i^j is to push unit i back to keep safe distance. By incorporating the driver's unsatisfied desire for mobility $(G_i - R_i)$, the net force in the x direction can be determined 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 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 k and the latter generates a correction force N_i if i deviates its lane center. Therefore, the net effect can be expressed as:

$$m_i \ddot{y}_i = \sum F_{i,y} = F_i^k - N_i = -\frac{\partial U_{i,y}}{\partial y}$$

By incorporating time t, unit i's perception-reaction time τ_i, and driver i's directional response γ, the above equations can be expressed as:

$$m_i \ddot{x}_i(t + \tau_i) = \sum \tilde{F}_{i,x}(t) = \gamma_i^0 [G_i(t) - R_i(t)] + \gamma(\alpha_i^j) \frac{\partial U_{i,x}}{\partial x}$$

$$m_i \ddot{y}_i(t + \tau_i) = \sum \tilde{F}_{i,y}(t) = -\gamma(\alpha_i^k) \frac{\partial U_{i,y}}{\partial y}$$

where $\gamma_i^0 \in [0, 1]$ represents the unit's attention to its unsatisfied desire for mobility (typically $\gamma_i^0 = 1$), α_i^j, α_i^k, and α_i^N are viewing angles which are also functions of time.

30.3.2 Microscopic modeling

The microscopic model can be formulated by simplifying the above picoscopic 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 roadway surfaces to a collection of lines.

Modeling longitudinal control

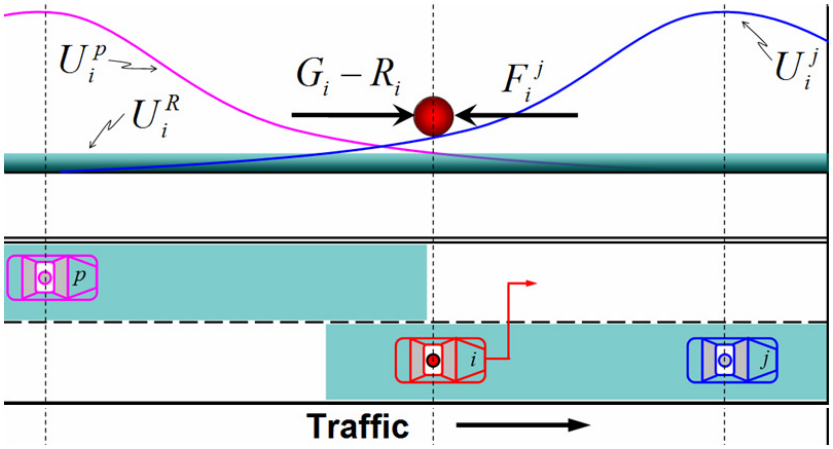


Figure 30.5: Microscopic modeling

With the above simplifications, the two-dimensional (3D) potential field U in Figure 30.4 reduces to a 2D potential function. The upper part of Figure 30.5 illustrates an example where a subject driver i (the middle one) is traveling behind a leading vehicle j and 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 of the figure 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 safe distance only comes from the leading vehicle and can be represented as:

$$F_i^j = -\frac{\partial U_i^j}{\partial x}$$

By incorporating roadway gravity G_i, roadway resistance R_i, and interaction between vehicles F_i^j, the net force on i can be expressed more specifically as:

$$m_i\ddot{x}_i = G_i - R_i - F_i^j$$

Departing from the above equation, a few special cases deserve particular attention:

$$\ddot{x}_i(t+\tau_i) = g_i[1 - (\frac{\dot{x}_i(t)}{v_i}) - e^{\frac{s_{ij}(t)^* - s_{ij}(t)}{s_{ij}(t)^*}}] \tag{30.1}$$

$$\ddot{x}_i(t+\tau_i) = g_i[1 - (\frac{\dot{x}_i(t)}{v_i}) - e^{\frac{s_{ij}(t)^* - s_{ij}(t)}{v_i\tau_i}}] \tag{30.2}$$

$$\ddot{x}_i(t+\tau_i) = g_i[1 - (\frac{\dot{x}_i(t)}{v_i}) - 2(1 - \frac{1}{1+e^{\frac{s_{ij}(t)^* - s_{ij}(t)}{v_i\tau_i/2+L_j}}})] \tag{30.3}$$

where it is assumed that $G_i = m_i \times g_i$, $R_i = m_i \times (\frac{\dot{x}_i(t)}{v_i})$, $F_i^j = m_i \times f(s_{ij}, s_{ij}(t)^*)$, g_i is the maximum acceleration that driver i is willing to apply when starting from stand still, $\dot{x}_i(t)$ 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, 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 the difference $(s_{ij}^* - s_{ij})$ represents how far vehicle i intrudes beyond s_{ij}^*. The rationale of representing the interaction force F_i^j between vehicles i and j using an exponential function is to set the desired spacing s_{ij}^* as a base line, beyond which the intrusion by unit i is translated exponentially to the repelling force acting on the unit. The desired spacing s_{ij}^* can be further determined as follows.

According to [46], the desired spacing 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 applies an emergency brake at rate $B_j > 0$. This rule results in

$$s_{ij}^*(t) = x_{i-1}(t) - x_i(t) \geq \frac{\dot{x}_i^2(t)}{2b_i} + \dot{x}_i\tau_i - \frac{\dot{x}_{i-1}^2(t)}{2B_{i-1}} + L_j$$

Alternatively, one may choose to set the desired spacing as a simplified function of the relative speed of the two vehicles:

$$s_{ij}^*(t) = L_j + L_j(\dot{x}_i(t) - \dot{x}_j(t))$$

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 30.5 can be interpreted as drivers j and p's personal spaces after accounting for lane barrier. 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, such 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.

30.3.3 Mesoscopic modeling

Mesoscopic modeling applies principles of Non-Equilibrium 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 $f(x, v, t)dxdv$ denotes the probability of having a vehicle within space range $(x, x+dx)$ and speed range $(v, v + dv)$ at time t (see Figure 30.6). 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 following a procedure similar to deriving the Boltzmann equation [143, 54] from basic principles. The classical Boltzmann equation describes particles moving in a 3D domain, so the first step is to reduce the 3D case to a 1D case which represents traffic moving on a unidirectional highway.

Existing models, in particular those based on Prigogine's work, are postulated. In order to derive the 1D Boltzmann equation from basic principles, a sound understanding of the mechanism of traffic evolution is required. Existing models, including a derived model [155, 74], assumed that the mechanism is vehicle "collision". For example, the fast follower i in the left panel of Figure 30.7) keeps its speed up to the collision point and then abruptly changes

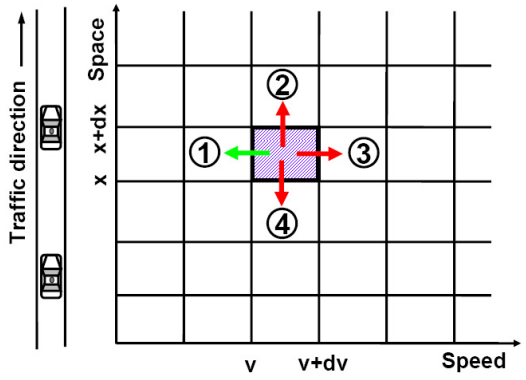


Figure 30.6: The x-v diagram

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 of Figure 30.7). This is possible only if car following is incorporated as the mechanism of particle interaction. As such, the longitudinal control model presented above can be used to derive the 1D Boltzmann equation and, thus, ensures micro-meso coupling.

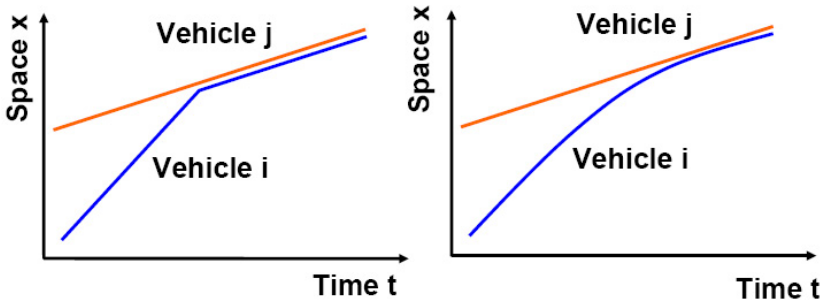


Figure 30.7: Car following

The derivation of the 1D Boltzmann equation starts by applying conservation law (e.g vehicles entering and exiting the highlighted cell in Figure 30.6 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) due to a sluggish leader, (2) vehicles

physically moved out of the cell, (3) vehicles accelerated by an aggressive follower, and (4) vehicles reversed, which is unlikely. The opposite applies to vehicles entering the cell. Therefore, applying the law to include all directions is the correct approach. Since deriving the 1D Boltzmann equation is mathematically complicated, this chapter only presents potential directions of exploration, leaving the actual derivation to be addressed in future research.

Once the 1D Boltzmann equation is formulated, one may solve it based on 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 under different temperature) which is the stationary (i.e. $\frac{\partial f}{\partial t} = 0$) solution to a classical Boltzmann equation.

30.3.4 Macroscopic modeling

Macroscopic modeling applies principles of Fluid Dynamics to model traffic flow as a 1D 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 $u(x,t) = \frac{1}{k}\int vf(x,v,t)dv$. Based on this understanding, it becomes clear that it is feasible to derive a hydrodynamical formulation from the mesoscopic model. The 1D 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 over v will give 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 $k(x,t)$. In order 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, obtained directly from the microscopic model by assuming equilibrium conditions. Presented below are a set of equilibrium v-k relationships derived from the special cases of the microscopic model, respectively:

$$v = v_f[1 - e^{1-\frac{k^*}{k}}] \tag{30.4}$$

$$v = v_f[1 - e^{1-\frac{1/k_j - 1/k}{v_f \tau}}] \tag{30.5}$$

$$v = v_f[\frac{2}{1 + e^{\frac{1/k_j - 1/k}{v_f \tau/2 + 1/k_j}}} - 1] \tag{30.6}$$

where $k^* = \tau v e^{-\frac{v}{v_f}} + \frac{1}{k_j}$, v_f is free-flow speed, $k_j = 1/L$, L is the bumper-to-bumper distance between vehicles when traffic is jammed, τ is 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.

$$\frac{\partial k}{\partial t} + \frac{\partial ku}{\partial x} = \int C dv$$
$$v = V(k)$$

The the system of equations can be solved using a finite difference method. A typical finite difference method is illustrated in Figure 30.8 where one partitions the time-space domain into cells and keeps track of traffic flowing into and out of each cell [115, 91, 19].

30.4 Conclusion and Future Directions

This chapter presents a broad perspective on traffic flow modeling at a spectrum of four scales: picoscopic, microscopic, mesoscopic, and macroscopic

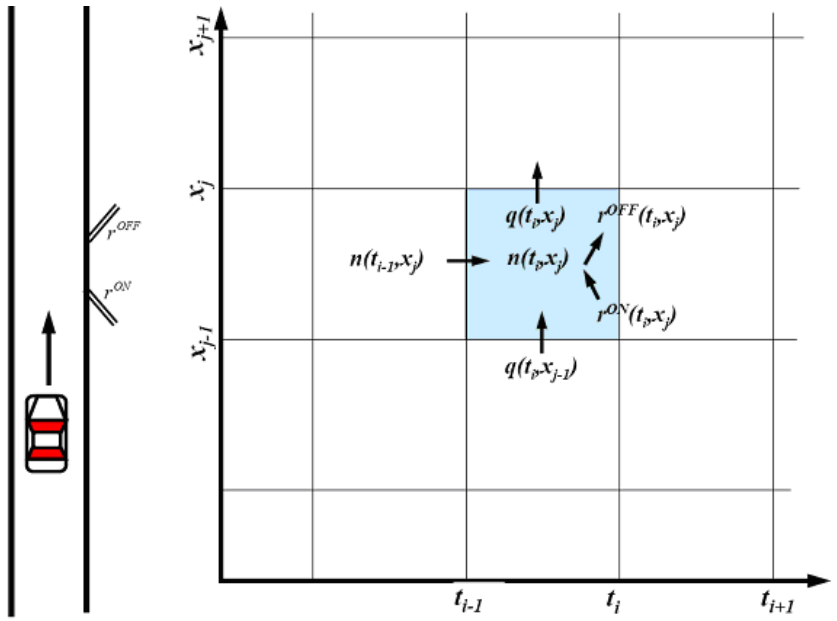


Figure 30.8: The finite difference method

from the most to the least detailed level in that order. Modeling objectives and model properties at each scale are discussed and existing efforts are reviewed.

In order to ensure modeling consistency and provide a microscopic basis for macroscopic models, it is critical to address the coupling among models at different scales, i.e. 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 is proposed based on field theory. Basically, in this approach, physical world 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 vehicle motion 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 are discussed. More specifically, the field theory serves as the basis of picoscopic modeling which represents a driver-vehicle unit as driver-vehicle-environment closed-loop control system. The system is able to capture vehicle motion in longitudinal and lateral directions. The microscopic model is obtained from the picoscopic model by simplifying its driver-vehicle interactions, vehicle dynamics, and vehicle lateral motion. The mesoscopic model is derived from basic principles using 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 derived from the microscopic model directly). Therefore, the proposed approach ensures model coupling and modeling consistency. As such, consistent models drived from this approach are able to provide the theoretical foundation to develop the "zoomable" traffic simulation tool discussed in Section 30.1.

A few special cases of the microscopic model are formulated. Further, their corresponding equilibrium speed-density relationships are derived. A numerical test is devised to verify if these microscopic special cases make any sense. In addition, the equilibrium relationships are validated against empirical data. Both numerical and empirical results suggest that these special cases perform satisfactorily and aggregate to realistic macroscopic behavior.

This chapter emphasizes modeling strategies at the four scales. Though a family of special cases are formulated at the microscopic and macroscopic scales, further efforts are needed to complete the spectrum by adding specific models at the picoscopic and mesoscopic scales. In addition, with the rapid development of wireless technologies and the deployment of connected vehicle technology (CVT), the effects of vehicle-vehicle and vehicle-roadside communications will transform the way a transportation system operates. Therefore, it is desirable that traffic flow models are able to incorporate such effects to realistically simulate CVT-enabled transportation systems.

Bibliography

[1] 2010 Annual Urban Mobility Report, Texas Transportation Institute (TTI). Accessible online at: `http://mobility.tamu.edu/ums/`, accessed on June 5, 2010.

[2] National Transportation Statistics, Bureau of Transportation Statistics (BTS). Accessible online at: `http://www.bts.gov/publications/national_transportation_statistics/`, accessed on June 5, 2010.

[3] Masato Abe. Theoretical Analysis on Vehicle Cornering Behaviours in Braking and in Acceleration. *Vehicle System Dynamics*, 14(1-3):140–143, 1985.

[4] K.I. Ahmed, M. Ben-Akiva, H.N. Koutsopoulos, and R.G. 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*, pages 501–515, 1996.

[5] B. Akinci, C. Hendrickson, and I. Karaesmen. Exploiting Motor Vehicle Information and Communications Technology for Transportation Engineering. *Journal of Transportation Engineering*, 129:469–474, 2003.

[6] M. D. Artamonov, V. A. Ilarionov, and M. M. Morin. *Motor Vehicles: Fundamentals and Design*. Mir Publishers, Moscow, 1976.

[7] G. Bar-Meir. *Fundamentals of Compressible Fluid Mechanics*. Potto Project, 2007.

[8] M. E. Ben-Akiva, M. Bierlaire, H. Koutsopoulos, and R. Mishalani. DynaMIT: A Simulation-Based System for Traffic Prediction. DACCORS Short Term Forecasting Workshop, 1998.

[9] R. Benekohal and J. Treiterer. Carsim: Car Following Model for Simulation of Traffic in Normal and Stop-and-Go Conditions. *Transportation Research Record*, 1194:99–111, 1988.

[10] K. L. Butler, M. Ehsani, and P. Kamath. Matlab-Based Modeling and Simulation Package for Electric and Hybrid Electric Vehicle Design. *IEEE Transactions on Vehicular Technology*, 48:1770–1778, 1999.

[11] R.E. Chandler, R. Herman, and E.W. Montroll. Traffic Dynamics: Studies in Car Following. *Operations Research*, 6:165–184, 1958.

[12] G.-L. Chang, T. Junchaya, and A. J. Santiago. A Real-Time Network Traffic Simulation Model for ATMS Applications: Part ISimulation Methodologies. *Journal of Intelligent Transportation Systems*, 1(3):227241, 1994.

[13] 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*, 215:495–506, 2001.

[14] D. Cho and J. K. Hedrick. Automotive Powertrain Modeling for Control. *Journal of Dynamic Systems, Measurement and Control, Transactions ASME*, 111:568–576, 1989.

[15] F. E. Coates and R. D. Fruechte. Dynamic Engine Models for Control Development. Part II: Application to Idle Speed Control. *International Journal of Vehicle Design SP4*, page 75, 1983.

[16] J. A. Cook and B. K. Powell. Modeling of an Internal Combustion Engine for Control Analysis. *IEEE Control Systems Magazine*, pages 20–26, 1988.

[17] P. R. Crossley and J. A. Cook. A Nonlinear Engine Model for Drivetrain System Development. In *International Conference on Control 1991, Vol. 2*, pages 921–925, Edinburgh , UK, 1991.

[18] C.F. Daganzo. The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory. *Transportation Research B*, 28(4):269–287, 1994.

[19] C.F. Daganzo. A Finite Difference Approximation of the Kinematic Wave Model of Traffic Flow. *Transportation Research B*, 29(4):261–276, 1995.

[20] C.F. Daganzo. Requiem For Second-order Fluid Approximations of Traffic Flow. *Transportation Research B*, 29(4):277–286, 1995.

[21] C.F. Daganzo. The Cell Transmission Mode, Part II: Network Traffic. *Transportation Research B*, 29(2):79–93, 1995.

[22] C. S. Daw, M. B. Kennel, C. E. A. Finney, and F. T. Connolly. Observing and Modeling Nonlinear Dynamics in an Internal Combustion Engine. *Journal of the American Physical Society: Review E*, 57:2811–2819, 1998.

[23] J. M. Del Castillo and F. G. Bentez. On the Functional Form of the Speed-Density Relationship - I: General Theory. *Transportation Research Part B: Methodological*, 29(5):373–389, 1995.

[24] J. M. Del Castillo and F. G. Bentez. On the Functional Form of the Speed-Density Relationship - II: Empirical Investigation . *Transportation Research Part B: Methodological*, 29(5):391–406, 1995.

[25] J. M. del Castillo, P. Pintado, and F. G. Benitez. A Formulation for the Reaction Time of Traffic Flow Models. In *Proceedings of the 12nd Int. Symp. Theory Traffic Flow, Paris, France*, pages 387–405, 1993.

[26] S. Delprat, J. Lauber, T. M. Guerra, and J. Rimaux. Control of a Parallel Hybrid Powertrain: Optimal Control. *IEEE Transactions on Vehicular Technology*, 53:872–881, 2004.

[27] D. J. Dobner. A Mathematical Engine Model for Development of Dynamic Engine Control. Technical report, Society of Automotive Engineers (SAE) Report No. 800054, 1980.

[28] D. J. Dobner. Dynamic Engine Models for Control Development. Part I: Nonlinear and Linear Model Formulation. *International Journal of Vehicle Design SP4*, pages 54–74, 1983.

[29] J.S. Drake, J.L. Schofer, and A.D. May. A Statistical Analysis of Speed Density Hypotheses. *Highway Research Record*, 154:53–87, 1967.

[30] D. R. Drew. *Traffic Flow Theory and Control*. McGraw-Hill Book Company, Chapter 12, 1968.

[31] D.R. Drew. Deterministic Aspects of Freeway Operations and Control. *Highway Research Record*, 99:48–58, 1965.

[32] D.R. Drew, L.R. LaMotte, J.H. Buhr, and J.A. Wattleworth. Gap Acceptance in the Freeway Merging Process. Technical report, Texas Transportation Institute, 430-432, 1967.

[33] L. C. Edie. Discussion on traffic stream measurements and definitions. In *Proceedings of the 2nd Int. Symp. Theory Traffic Flow, Paris, France*, page 139154, 1963.

[34] L.C. Edie. Car-Following and Steady State Theory for Noncongested Traffic. *Operations Research*, 9, 1961.

[35] M. Ehsani, Y. M. Gao, S. E. Gay, and A. Emadi. *Modern Electric, Hybrid Electric, and Fuel Cell Vehicles*. CRC Press, 2005.

[36] FMCSA. FMCSA Part 571: Federal Motor Vehicle Safety Standards: Standard No. 105; Hydraulic and Electric Brake Systems. `http://www.fmcsa.dot.gov/`, 1999.

[37] T. Forbes and M. Simpson. Driver and Vehicle Response in Freeway Deceleration Waves. *Transportation Science*, 2(1):77–104, 1968.

[38] T.W. Forbes. Human Factor Considerations in Traffic Flow Theory. *Highway Research Record*, 15:60–66, 1963.

[39] T.W. Forbes, H.J. Zagorski, E.L. Holshouser, and W A. Deterline. Measurement of Driver Reaction to Tunnel Conditions. *Proceedings of Highway Research Board*, 37:345–357, 1958.

[40] D. W. Gao, C. Mi, and A. Emadi. Modeling and Simulation of Electric and Hybrid Vehicles. In *Proceedings of the IEEE, Vol. 95*, pages 729–745, 2007.

[41] Zhenhai Gao, Nanning Zheng, Hsin Guan, and Konghui Guo. Application of Driver Direction Control Model in Intelligent Vehicle's Decision and Control Algorithm. In *Intelligent Vehicle Symposium, 2002. IEEE*, volume 2, pages 413–418, 2002.

[42] Nathan Gartner, Carroll J. Messer, and Ajay K. Rathi. *Revised Monograph of Traffic Flow Theory: A State-of-the-Art Report*. Transportation Research Board, 2001.

[43] D. C. Gazis, R. Herman, and R. W. Rothery. Non-Linear Follow the Leader Models of Traffic Flow. *Operations Research*, 9:545–567, 1961.

[44] G. Genta. *Motor Vehicle Dynamics: Modeling and Simulation*. World Scientific, 2003.

[45] T. D. Gillespie. *Fundamentals of Vehicle Dynamics*. Society of Automotive Engineers, Inc., 1992.

[46] P.G. Gipps. A Behavioral Car Following Model for Computer Simulation. *Transportation Research, Part B*, 15:105–111, 1981.

[47] P.G. Gipps. A Model for the Structure of Lane-Changing Decisions. *Transportation Research, Part B*, 20:403–414, 1986.

[48] T.J. Gordon and M.C. Best. On the Synthesis of Driver Inputs for the Simulation of Closed-Loop Handling Manoeuvres. *International Journal of Vehicle Design*, 40(1-3):52–76, 2006.

[49] H. Greenberg. An Analysis of Traffic Flow. *Operations Research*, 7:78–85, 1959.

[50] B.D. Greenshields. A study of Traffic Capacity. *Proceedings of Highway Research Board*, 14:448–477, 1934.

[51] D. J. Grieve. Simulation of Dynamic Systems - MECH 337. `http://www.tech.plym.ac.uk/sme/mech331/partb1.htm`, Accessed November 20, 2006.

[52] F.A. Haight. Vehicles as Particles. (Book Reviews: Kinetic Theory of Vehicular Traffic by Prigogine and Herman). *Science*, 173(3996):513, 1971.

[53] M.M. Hamed, S.M. Sama, and R.R. Batayneh. Disaggregate Gap-Acceptance Model for Unsignalised T-Intersections. *Journal of Transportation Engineering, ASCE*, 123(1):36–42, 1997.

[54] Stewart Harris. *An Introduction to the Theory of the Boltzmann Equation*. Dover Books on Physics. Dover Publications, 2004.

[55] D. Helbing, A. Hennecke, V. Shvetsov, and M. Treiber. Micro- and Macro-Simulation of Freeway Traffic. *Mathematical and Computer Modelling*, 35(5):517–547, 2002.

[56] Dirk Helbing. Theoretical Foundation of Macroscopic Traffic Models. *Physica A: Statistical and Theoretical Physics*, 219(3-4):375–390, 1995.

[57] Dirk Helbing. Traffic and Related Self-Driven Many-Particle Systems. *Reviews of Modern Physics*, 73:1067–1141, 2001.

[58] Dwayne Henclewood and Daiheng Ni. A Dynamic-Interactive-Vehicle Model for Modeling Traffic beyond the Microscopic Level. *International Journal of Vehicle Information and Communication Systems (IJVICS)*, 2(1-2), 2009.

[59] R. Herman and G.H. Weiss. Comments on the Highway Crossing Problem. *Operations Research*, 9:838–840, 1961.

[60] R.H. Hewitt. Using Probit Analysis with Gap Acceptance Data. Technical report, Department of Civil Enginneering, University of Glasgow, 1992.

[61] J. B. Heywood. *Internal combustion Engine Fundementals*. McGraw-Hill, New York, 1989.

[62] P. Hidas and 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*, pages 53–69, 1999.

[63] C.W. Hong. Automotive Dynamic Performance Simulator for Vehicular Powertrain System Design. *International Journal of Vehicle Design*, 16:264–281, 1995.

[64] S.P. Hoogendoorn and P.H.L. Bovy. State-of-the-art of Vehicular Traffic Flow Modelling. *Proceedings of the IMechE Part I, Journal of Systems and Control Engineering*, 215(4):283–303, 2001.

[65] V. F. Hurdle and B. Son. Road Test of a Freeway Model. *Transportation Research, Part A*, 34(7):537–564, 2000.

[66] R. Illner, C. S. Bohun, S. McCollum, and T. V. Roode. *Mathematical Modelling: A Case Study Approach*. American Mathematical Society, 2005.

[67] Marita Irmscher, Thomas Jurgensohn, and Hans-Peter Willumeit. Driver Models in Vehicle Development. *Vehicle System Dynamics*, 33(Suppl):83–93, 1999.

[68] Hongfei Jia, Zhicai Juan, and Anning Ni. Develop a car-following model using data collected by "five-wheel system". In *The Proceedings of the 2003 IEEE Intemational Confuence on Intelligent Transportation Systems*, volume 1, pages 346–351, 2003.

[69] M. Kabganian and R. Kazemi. A New Strategy for Traction Control in Turning via Engine Modeling. *IEEE Transactions on Vehicular Technology*, 50:1540–1548, 2001.

[70] B.S. Kerner and P. Konhauser. Cluster Effect in Initially homogeneous Traffic Flow. *Physical Review E*, 48(4):2335–2338, 1993.

[71] R.D. Khne. Macroscopic Freeway Model for Dense Traffic - Stop-Start Waves and Incident Detection. In *Ninth International Symposium on Transportation and Traffic Theory*, pages 20–42, 1984.

[72] R.D. Khne. Freeway Control and Incident Detection Using a Stochastic Continuum Theory of Traffic Flow. In *1st International Conference on Applied Advanced Technology in Transportation Engineering*, pages 287–292, San Diego, CA, 1989.

[73] S. Kikuchi and P. Chakroborty. Car-Following Model Based on a Fuzzy Inference System. *Transportation Research Record*, 1365:82–91, 1992.

[74] A. Klar and R. Wegener. A Hierarchy of Models for Multilane Vehicular Traffic (Part I: Modeling and Part II: Numerical and Stochastic Investigations). *SIAM Journal on Applied Mathematics (SIAP)*, 59:983–1011, 1999.

[75] Roger Knobel. *An Introduction to the Mathematical Theory of Waves*. American Mathematical Society, 2000.

[76] I. Kosonen. *HUTSIM - Urban Traffic Simulation and Control Model: Principles and Applications*. PhD thesis, Helsinki University of Technology, 1999.

[77] U. Kramer and G. Rohr. A Model of Driver Behaviour. *Ergonomics*, 25(10):891–907, 1982.

[78] C. J. Leo and R. L. Pretty. Numerical Simulation of Macroscopic Continuum Traffic Models. *Transportation Research, Part B*, 26(3):207–220, 1992.

[79] Jia Li, Haizhong Wang, Qian-Yong Chen, and Daiheng Ni. Analysis of LWR Model with Fundamental Diagram Subject to Uncertainties. *To appear in Transportmetrica, Taylor and Francis*, 2011.

[80] M. Lighthill and G. Whitham. On Kinematic Waves II. A Theory of Traffic Flow on Long Crowded Roads. *Proc. Royal Society of London, Part A*, 229(1178):317–345, 1955.

[81] Y. Lin, P. Tang, W.J. Zhang, and Q. Yu. Artificial Neural Network Modelling of Driver Handling Behaviour in a Driver-Vehicle-Environment System. *International Journal of Vehicle Design*, 37(1):24–45, 2005.

[82] Erik M. Lowndes and J.W. David. Development of an Intermediate Degree of Freedom Vehicle Dynamics Model for Optimal Design Studies. *American Society of Mechanical Engineers, Design Engineering Division (Publication) DE*, 106:19–24, 2000.

[83] Z. D. Ma and N. C. Perkins. An Efficient Multibody Dynamics Model for Internal Combustion Engine Systems. *Multibody System Dynamics*, 10:363–391, 2003.

[84] Charles C. Macadam. Understanding and Modeling the Human Driver. *Vehicle System Dynamics*, 40(1-3):101–134, 2003.

[85] Charles C. Macadam and Gregory E. Johnson. Application of Elementary Neural Networks and Preview Sensors for Representing Driver

Steering Control Behaviour. *Vehicle System Dynamics*, 25(1):3–30, 1996.

[86] Y. Makigami, G. Newell, and R. Rothery. Three-Dimensional Representation of Traffic Flow. *Transportation Science*, 5(3):302313, 1971.

[87] A.D. May. FREQ User Manual. Technical report, California Department of Transportation, Berkeley, 1998.

[88] Adolf D. May. *Traffic Flow Fundamentals*. Prentice-Hall, 1989.

[89] 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.

[90] P. G. Michalopoulos, J. K. Lin, and D. E. Beskos. Integrated Modelling and Numerical Treatment of Freeway Flow. *Appl. Mathem. Model*, 11(401):447–458, 1987.

[91] P.G. Michalopoulos. Dynamic Freeway Simulation Program for Personal Computers. *Transportation Research Record*, 971:68–79, 1984.

[92] P.G. Michalopoulos, P. Yi, and A.S. Lyrintzis. Continuum modelling of Traffic Dynamics for Congested Freeways. *Transportation Research B*, 27:315–332, 1993.

[93] 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, September 2003.

[94] R. G. Mortimer, L. Segel, H. Dugoff, J. D. Campbell, C. M. Jorgeson, and 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. Technical report, Highway Safety Research Institute, Ann Arbor, Mich., 1970.

[95] J. J. Moskwa. *Automotive Engine Modeling for Real Time Control*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1988.

[96] J.J. Moskwa and J.K. Hedrick. Automotive Engine Modeling for Real Time Control Application. In *Proceedings of the 1987 American Control Conference*, pages 341–346, Minneapolis, MN, 1987.

[97] P.K. Munjal and L.A. Pipes. Propogation of On-Ramp Density Perturbations on Uni-Directional and Two- and Three-Lane Freeways. *Transportation Research Part B*, 5(4):241–255, 1971.

[98] P. Nelson. A Kinetic Model of Vehicular Traffic and its Associated Bimodal Equilibrium Solutions. *Transport Theory and Statistical Physics*, 24:383–409, 1995.

[99] G. F. Newell. Nonlinear Effects in the Dynamics of Car Following. *Operations Research*, 9(2):209–229, 1961.

[100] G.F. Newell. A Simplified Theory on Kinematic Waves in Highway Traffic, Part I: General Theory. *Transportation Research B*, 27(4):281–287, 1993.

[101] G.F. Newell. A Simplified Theory on Kinematic Waves in Highway Traffic, Part II: Queueing at Freeway Bottlenecks. *Transportation Research B*, 27(4):289–303, 1993b.

[102] G.F. Newell. A Simplified Theory on Kinematic Waves in Highway Traffic, Part III: Multi-Destination Flows. *Transportation Research B*, 27(4):305–313, 1993c.

[103] G.F. Newell. A Simple Car-Following Theory: A Lower Order Model. *Transportation Research, Part B*, 36(3):195–205, 2002.

[104] Daiheng Ni. *Extension and Generalization of Newell's Simplified Theory of Kinematic Waves*. PhD thesis, Department of Civil and Environmental Engineering, Georgia Institute of Technology, 2004.

[105] Daiheng Ni. Challenges and Strategies of Transportation Modeling and Simulation under Extreme Conditions. *International Journal of Emergency Management (IJEM)*, 3(4):298–312, 2006.

[106] Daiheng Ni. Determining Traffic Flow Characteristics by Definition for Application in ITS. *IEEE Transactions on Intelligent Transportation Systems*, 8(2):181–187, 2007.

[107] Daiheng Ni. Multiscale Modeling of Traffic Flow. *Mathematica Aeterna, Hilaris Ltd.*, 1(01):27–54, 2011.

[108] Daiheng Ni. A Framework for New Generation Transportation Simulation. In *Proceedings of Winter Simulation Conference '06*, Portola Plaza Hotel, Monterey, CA, December 3-6, 2006.

[109] Daiheng Ni and Dwayne Henclewood. Simple Engine Models for VII-Enabled In-Vehicle Applications. *IEEE Transactions on Vehicular Technology*, 57(5):2695–2702, 2008.

[110] Daiheng Ni, John D. Leonard, and Billy M. Williams. The Network Kinematic Waves Model: A Simplified Approach to Network Traffic. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations.*, 10(1):1–14, 2006.

[111] S. Panwai and H. F. Dia. A Reactive Agent-based Neural Network Car Following Model. In *8th International IEEE Conference on Interlligent Transportation Systems (ITSC2005)*, pages 326–331, Vienna, Austria, 2005.

[112] Sakda Panwai and Hussein Dia. Neural Agent Car-Following Models. *IEEE Transactions on Intelligent Transportation Systems*, 8(1):60–70, 2007.

[113] S.L. Paveri-Fontana. On Boltzmann-like Treatments for Traffic Flow: A Critical Review of the Basic Model and an Alternative Proposal for Dilute Traffic Analysis. *Transportation Research*, 9:225–235, 1975.

[114] H.J. Payne. Models of Freeway Traffic and Control. In *Simulation Council Proceedings*, volume 1, pages 51–61, 1971.

[115] H.J. Payne. FREFLO: A Macroscopic Simulation Model for Freeway Traffic. *Transportation Research Record*, 722:68–77, 1979.

[116] K. Petty. Small Time Scale Analysis of Loop Data. `http://ipa.eecs.berkeley.edu/~pettyk/FSP/`, Accessed November 20, 2006.

[117] W.F. Phillips. Kinetic Model for Traffic Flow. Technical report, Report DOT/RSPD/DPB/50-77/17. U. S. Department of Transportation, 1977.

[118] W.F Phillips. A Kinetic Model for Traffic Flow with Continuum Implications. *Transportation Planning and Technology*, 5:131–138, 1979.

[119] L. A Pipes. An Operational Analysis of Traffic Dynamics. *Journal of Applied Physics*, 24:271–281, 1953.

[120] L.A. Pipes. Car Following Models and Fundamental Diagram of Road Traffic. *Transportation Research*, 1, 1967.

[121] B. K. Powell. A Dynamic Model for Automotive Engine Control Analysis. In *18th IEEE Decision and Control Conference*, 1979.

[122] 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.

[123] I. Prigogine. *A Boltzmann-like Approach to the Statistical Theory of Traffic Flow.* Theory of traffic Flow. Elsevier, Amsterdam, 1961.

[124] I. Prigogine and F. C. Andrews. A Boltzmann-like Approach for Traffic Flow. *Operations Research*, 8(6), 1960.

[125] P. F. Puleston, G. Monsees, and 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(117-124), 2002.

[126] M.S. Raff and J.W. Hart. A Volume Warrant for Urban Stop Signs. Technical report, Eno Foundation for Highway Traffic Control, Saugatuck, Connecticut, 1950.

[127] H. Rakha and I. Lucic. Variable Power Vehicle Dynamics Model for Estimating Truck Acceleration. *Journal of Transportation Engineering*, 128(5):412–419, 2002.

[128] H. Rakha, I. Lucic, S. H. Demarchi, J. R. Setti, and M. V. Aerde. Vehicle Dynamics Model for Predicting Maximum Truck Acceleration Levels. *Journal of Transportation Engineering*, 127(5):418–425, 2001.

[129] H. Rakha, M. Snare, and F. Dion. Vehicle Dynamics Model for Estimating Maximum Light Duty Vehicle Acceleration Levels. *Transportation Research Record*, 1883(1):40–49, 2004.

[130] P.I. Richards. Shock Waves on the Highway. *Operations Research*, 4:42–51, 1956.

[131] G. Rizzoni. Estimate of Indicated Torque from Crankshaft Speed Fluctuations: A Model for The Dynamics of the IC Engine. *IEEE Transactions on Vehicular Technology*, 38:168–179, 1989.

[132] G. Rizzoni. Stochastic Model for the Indicated Pressure Process and the Dynamics of the Internal Combustion Engine. *IEEE Transactions on Vehicular Technology*, 38:180–192, 1989.

[133] Riccardo Rossi and Claudio Meneguzzer. 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.

[134] J. J. Scillieri, J. H. Buckland, and J. S. Freudenberg. Reference Feedforward in the Idle Speed Control of a Direct-Injection Spark-Ignition Engine. *IEEE Transactions on Vehicular Technology*, 54:51–61, 2005.

[135] Y. Shiao, C. H. Pan, , and J. J. Moskwa. Advanced dynamic Spark Ignition Engine Modelling for Diagnostics and Control. *International Journal of Vehicle Design*, 15:578–596, 1994.

[136] L. Smith, R. Beckman, D. Anson, K. Nagel, and M.E. Williams. TRANSIMS: Transportation Analysis and Simulation System. In *Fifth National Conference on Transportation Planning Methods Applications-Volume II*, 1995.

[137] S A Smith. Freeway Data Collection for Studying Vehicle Interactions - Technical Report. Final Report. Technical report, JHK & Associates and Federal Highway Administration, 1985.

[138] B. Son. *A Study of G.F. Newell's Simplified Theory of Kinematic Waves in Highway Traffic*. PhD thesis, Department of Civil Engineering, University of Toronto, Canada, 1996.

[139] A. D St. John. Grade Effects on Traffic Flow Stability And Capacity. Technical report, National Cooperative Highway Research Report 185. Transportation Research Board, National Research Council, 1978.

[140] Rahul Sukthankar. *Situation Awareness for Tactical Driving*. PhD thesis, Carnegie Mellon University, 1997.

[141] D. Swaroop, J. K. Hedrick, C. C. Chien, and P. Ioannou. Comparision of Spacing and Headway Control Laws for Automatically Controlled Vehicles. *Vehicle System Dynamics*, 23:597–625, 1994.

[142] M. Szeto and D. Gazis. Application of Kalman Filtering to the Surveillance and Control of Traffic Systems. *Transportation Science*, 6(4):419439, 1972.

[143] Richard C. Tolman. *The Principles of Statistical Mechanics*. Dover Publications, 1980.

[144] M. Treiber, A. Hennecke, and D. Helbing. Derivation, Properties and Simulation of A Gas-kinetic-based, Nonlocal Traffic Model. *Physical Review E*, 59(1):239–253, 1999.

[145] Martin Treiber, Ansgar Hennecke, , and Dirk Helbing. Congested Traffic States in Empirical Observations and Microscopic Simulations. *Phys. Rev. E*, 62:18051824, 2000.

[146] R.T. Underwood. Speed, Volume and Density Relationships. *Quality and Theory of Traffic Flow, Yale University Report, New Haven, Connecticut*, 1961.

[147] 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, D.C., 1995.

[148] M. Van Aerde and H. Rakha. Multivariate Calibration of Single Regime Speed-Flow-Density Relationships. In *Proceedings of the 6th 1995 Vehicle Navigation and Information Systems Conference*, pages 334–341, Seattle, WA, USA, 1995.

[149] S. M. Velan and M. Van Aerde. Gap Acceptance and Approach Capacity at Unsignalized Intersections. *ITE Journal*, 66(3):40–45, 1996.

[150] J. R. Wagner, D. M. Dawson, and L. Zeyu. Nonlinear Air-to-Fuel Ratio and Engine Speed Control for Hybrid Vehicles. *IEEE Transactions on Vehicular Technology*, 52:184–195, 2003.

[151] Haizhong Wang, Jia Li, Qian-Yong Chen, and Daiheng 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.

[152] Haizhong Wang, Jia Li, Qian-Yong Chen, and Daiheng Ni. Logistic Modeling of the Equilibrium Speed-Density Relationship. *Transportation Research Part A, Elsevier*, 45(6):554566, 2011.

[153] Haizhong Wang, Daiheng Ni, Qian-Yong Chen, and Jia Li. Stochastic Modeling of Equilibrium Speed-Density Relationship. *To appear in Journal of Advanced Transportation, John Wiley & Sons*, 2011.

[154] J. G. Wardrop. Some Theoretical Aspects of Road Traffic Research. In *Proceedings of the Institution of Civil Engineers, Part II, Volume I*, pages 325–362, 1952.

[155] R. Wegener and A. Klar. A Kinetic Model for Vehicular Traffic Derived from a Stochastic Microscopic Model. *Transport Theory and Statistical Physics*, 25:785–798, 1996.

[156] G.B. Whitham. *Linear and Nonlinear Waves*. John Wileyand Sons Inc, New York, NY., 1974.

[157] R. Wiedemann. *Simulation des Straenverkehrsflusses*. PhD thesis, Schriftenreihe des Instituts fr Verkehrswesen der Universitt Karlsruhe, Germany, 1974.

[158] W.W. Wierwille, G.A. Gagne, and J.R. Knight. An Experimental Study of Human Operator Models and Closed-Loop Analysis Methods for High-Speed Automobile Driving. *IEEE Transactions on Human Factors in Electronics*, HFE-8(3):187–201, 1967.

[159] S. Yager. CORQ - A Model for Predicting Flows and Queues in a Road Corridor. *Transportation Research Record*, 533:77–87, 1975.

[160] J. Yi, H. Lin, L. Alvarez, and R. Horowitz. Stability of Macroscopic Traffic Flow Modeling Through Wavefront Expansion. *Transportation Research, Part B*, 37(7):661–679, 2003.

[161] P. Yoon and M. Sunwoo. A Nonlinear Dynamic Modelling of SI Engines for Controller Design. *International Journal of Vehicle Design*, 26:277–297, 2001.

[162] A.G. Zadeh, A. Fahim, and M. El-Gindy. Neural Network and Fuzzy Logic Applications to Vehicle Systems: Literature Survey. *International Journal of Vehicle Design*, 18(2):132–193, 1997.

[163] H.M. Zhang. A Theory of Nonequilibrium Traffic Flow. *Transportation Research B*, 32(7):485–498, 1998.

[164] Y. Zhang, L.E. Owen, and J.E. Clark. A Multi-regime Approach for Microscopic Traffic Simulation. In *77th Transportation Research Board Annual Meeting*, 1998.